University of Sussex
Economics

Working Paper Series

No. 88-2016

Disciplinary Sanction and Social Pressure in English Premiership Soccer

Barry Reilly[1] and Robert Witt[2]

1.Department of Economics, University of Sussex, Brighton, BN1 9RF UK

b.m.reilly@sussex.ac.uk

2. School of Economics, University of Surrey, Guildford, GU2 7XH, UK

r.witt@surrey.ac.uk

Abstract: This paper uses player/match level data drawn from five playing seasons of the English Premiership League (EPL) to test for the presence of a refereeing susceptibility to social pressure in the application of soccer's commonest sanction, the yellow disciplinary card. Using both player-specific fixed and random effects models, tentative support for the proposition is uncovered. The estimated effect, however, is found to be negligible in magnitude and unlikely to influence match outcomes in a meaningful way.

JEL classification: C23, D81, L83

Key words: Disciplinary Sanction, Referees, Social pressure

The authors are grateful to the Nuffield Foundation for financial support (under award SGS/36059). The excellent research assistance provided by Edgar Cooke and Kalle Hirvonen is acknowledged.

Introduction

Economists have recently exploited data from the highest tiers of professional soccer in Europe to empirically test whether social pressure influences a referee's intended impartial behaviour. The approach is motivated in part by a theoretical literature that links individual behaviour to the social environment (e.g., see Akerlof (1980), Bernheim (1994) and Becker and Murphy (2000)). Garicano *et al*. (2005) provided the first empirical study to exploit this conceptual framework and using data from Spain's La Liga reported evidence of a referee bias favouring the home team. The authors interpreted this as attributable to the influence of social pressure. The proxy for social pressure used in their study was stadium attendance and the measure for bias was the length of injury time awarded in close games.[1]

The use of 'added time' in professional soccer as a measure to inform referee bias is subject to criticism (see Rickman and Witt (2008)). For instance, the efforts of a home side losing in a tightly contested match may be undermined by the time-wasting tactics of their opponents prompting a referee to add more than the minimum time announced to compensate. In addition, the 'added time' reported in the published records on football matches may not be accurate raising the possibility of measurement error in the dependent variable. The limitations associated with this measure of bias encouraged the use of alternatives. For example, Sutter and Kocher (2004) chose the award of penalties in the German Bundesliga, Dohman (2008) used penalties, goals and 'added time' from the same league, Dawson and Dobson (2010) opted for disciplinary cards in their international study of European Cup matches, Pettersson-Lidbom and Priks (2010), using Italian data from a natural experiment, favoured fouls and disciplinary cards[2], while Reilly and Witt (2013) used the harshest sanction available to referees (i.e., red cards) as the outcome of interest for their study of the English Premiership.

The empirical evidence reported in most of these studies generally suggests the presence of a referee home bias animated by social pressure. The use of a variety of alternative outcomes to measure referee bias enhances the research agenda for this topic and also provides some useful robustness checks. However, the majority of studies cited suffer from other possible deficiencies that merit caution in interpreting the reported evidence. First, all the aforementioned papers with the exception of Reilly and Witt (2013) use match-level data, which preclude the role of player heterogeneity, team attributes and game-specific characteristics from determining the selected outcomes. These omissions may be particularly serious in the context of penalty and sanctioning decisions. Second, all these studies, with the exceptions of Dawson and Dobson (2010) and Reilly and Witt (2013), use the ordinary least squares (OLS) procedure though the outcomes modelled are invariably truncated at zero for some (or even many) observations. For instance, the match-level summary statistics reported in both Garicano *et al*. (2005) and Rickman and Witt (2008) reveal zero values for their 'added time' variable, but in neither case is the scale of the censorship reported. The data on penalties and goals used by Dohman (2008) are likely to suffer from a similar type of censorship, as are the disciplinary outcomes favoured by Pettersson-Lidbom and Priks (2010).

Referees in professional soccer are generally required to make split-second decisions that potentially permit the influence of social pressure. The case appears to be particularly apposite with regard to the use of formal disciplinary sanctions within a match (see Pettersson-Lidbom and Priks (2010)). The social pressure effect is assumed mediated through stadium crowd noise, the level of which generally rises in response to the perceived foul play of visiting team players. Foul play and other miscreant behaviour on the part of players can result in the award of a yellow disciplinary card, with the accumulation of two such cards (i.e., a red card) resulting in dismissal from the field of play. In certain (and rarer) circumstances, players can be shown a straight red card for offences deemed to constitute excessively dangerous play. The use of this latter sanction weakens a team numerically and Ridder *et al*. (1994), using data from professional soccer leagues in the Netherlands, highlight the potential effect on a match outcome of a player's dismissal.[3] The use of the lesser of the two sanctions could also weaken the team as yellow carded players may adopt a more cautious approach in their tackling and overall general play through fear of incurring a second sanction culminating in their expulsion from the match. As noted by Dawson *et al*. (2007), the spectre of the referee brandishing a disciplinary card of any colour to an offending player enhances the drama of a football match. In addition, the theatrical action of carding a visiting team player may also serve to appease the majority of those in attendance, thus allowing a referee to gain favour with the home crowd.

In contrast to referees who officiate in the top tiers of professional soccer in Spain, Italy and Germany, those operating in the English Premiership League (EPL) over the period covered by our analysis are professional. In addition, they are subject to a very high degree of scrutiny that should potentially attenuate the influence of social pressure on their actions. They are monitored at each game by a match assessor who grades their match day performance and are required to attend fortnightly meetings at which their performances are evaluated.[4] Furthermore, referees are also required to operate under the glare of the print and broadcast media with televised Premiership games in particular ensuring that decisions of referees are subject to a fairly thorough degree of post-match analysis.

The exercise of a home bias along the player disciplining dimension may have more serious implications for a team than the award of a few extra minutes at the end of a match, the measure used by Garicano *et al*. (2005) and Rickman and Witt (2008). Specifically, using data from the EPL over five seasons, we examine the evidence of a referee home bias in the application of the yellow card sanction and assess the extent to which match attendance, the proxy used in this and other studies for social pressure, influences the disciplining behaviour of referees. The EPL provides a suitable focus for the interrogation of this research question. The teams participating in the league are generally competitive and the stadia, in contrast to some in the

German Bundesliga for example, do not contain athletics running tracks that potentially mitigate crowd noise levels. In contrast to Italy and Germany, none of the stadia in the EPL are shared by competing teams. Furthermore, the outcome of professional soccer games in England has not been tainted by corrupt refereeing practices, as recently the case in Italy (see Boeri and Severigini (2011)).

An important contribution of this study is its use of richer player/match level information than exploited to date in the empirical literature on this topic, which has almost exclusively relied on match-level data.[5] We take the view that in using a player disciplining framework to investigate the role of social pressure, the use of player-level data is more appropriate than aggregate match-level data, the latter of which may conceal the role of important heterogeneous factors. In particular, the use of information at the level of the player over games facilitates the explicit control of heterogeneous unobservable player effects that may potentially influence the award of a disciplinary card. Further, the econometric modelling adopted explicitly addresses the limited (or binary) nature of the dependent variable used in this application.

The organization of the paper is now outlined. The next section discusses the dataset to be used and is followed by a section that briefly outlines the econometric methodology. The empirical results are contained in the paper's penultimate section and a final one offers some concluding remarks.

Data

The data are obtained from OPTA *Sportsdata* and relate to games played in the EPL over five playing seasons between 2003/4 and 2007/8 inclusive. The data provide the club affiliation for each player, their age, primary field position, and number of club appearances. The criterion for the inclusion of players in the analysis requires a player to have made at least one appearance of whatever duration in any Premiership league game over the relevant period. The data also contain information on sanctions applied and include whether or not a player was shown a yellow card in a given match. A number of match-level variables are also available including the identity of the officiating referee, whether the player completed the full 90 (or more) minutes of play, whether or not the match was a ‘derby’ game, whether it was a home fixture for the player, and the size of match attendance. It is these latter two variables that provide the basis for our investigation of social pressure effects.

We have also constructed some other time-varying covariates including the number of games played in the season prior to the current fixture, and whether the player was over 32 years of age at the start of the relevant season, the definition we use here for a ‘veteran’ player. The former variable is included to reflect either a player’s recent form or fitness but could also reflect player learning in regard to disciplining which is enhanced with the number of games played. The latter variable potentially proxies for a player’s experience and guile in avoiding a sanction but, conversely, could also capture declining performance levels that may render older players more prone to receiving a sanction. We also include a variable for the difference in league positions between the two teams at the start of the game to reflect the degree of competitiveness of the fixture. Dummy variables for the 31 referees and the 30 clubs who featured in the Premiership over the seasons covered by our data are also included in the analysis. The inclusion of the latter set of variables is designed to control for the different playing styles of teams, with more aggressive styles likely to elicit a harsher sanctioning response from referees. Unfortunately, we cannot disaggregate the attendance figure into ‘home’ and ‘away’ support though traditionally in the EPL the latter represents a small fraction of those in attendance. In order to control for possible trend effects in a match official’s use of the yellow card, a set of year dummies reflecting the season in which the game was played is also included in the empirical specifications.

It is recognised that there is likely to be a causal relationship between fouls called and whether a yellow card is shown or not. Although our data identify the time a yellow card was issued within a particular game, they do not provide specific information on when the fouls were committed in the match. Thus, we are not in position to determine the foul count prior to a card being issued and so, consequently, the number of fouls called does not feature in the empirical modelling. However, we believe this particular limitation is mitigated by the inclusion of player-specific effects.

The resultant data comprise an unbalanced panel of 51,076 player/match-level observations on 1,162 players. The average number of observations (or matches) available per player is 44, with a minimum of 1 for a very small number of players and a maximum of 188. Table 1 provides a description of the variables with summary statistics. The table also separates the data by 'home' and 'away' fixture status. It reveals that the average yellow card rate per player per match is just over 11%. The 'home' players are found to have a 9.6% chance of receiving a yellow card, while 'away' players are subject to almost a 13% chance. Thus, 'away' team players are a third more likely to be shown a yellow card, on average. Although explanations for the observed differential are diverse, Boyco *et al*. (2007) suggest it partly explains the average match outcome advantage home sides enjoy in the EPL.

Table 1 about here

Econometric Methodology

Two different estimators are used to model the probability that player i receives a yellow card in match j within season t. The player provides the unit of observation and the panel data structure, as described in the previous section, permits an explicit treatment of unobservable player-specific effects. Given the binary nature of the dependent variable, we use two alternative variants of the logit model incorporating fixed and random effects respectively. A conventional binary logistic model could be augmented by including the full set of player-specific fixed effects (or dummies) but the conventional maximum likelihood (ML) estimation procedure in this case suffers from the well-known 'incidental parameters' problem and yields inconsistent parameter estimates. Chamberlain (1980) suggests maximizing a log-likelihood function conditional on a set of minimal sufficient statistics designed to sweep out the fixed effects. The model parameters are identified through the 'within' dimension of the data and, as with the linear regression fixed effects model, only the effects of time-varying covariates are identifiable.

A random effects logistic model is also used as an alternative to Chamberlain's model with the omitted effects assumed to be Gaussian distributed.[6] The treatment of unobserved player effects as random rather than fixed may appear less plausible in the current application since close to the full population of first team English Premiership footballers is used in our analysis. However, it is ultimately an empirical question whether the fixed or random effects model is more suitable in a particular application. A Hausman-type test can be used to test for the appropriateness of the random relative to the fixed effects model. The test statistic in its matrix form is expressed as:

$$[\boldsymbol{\beta}_{\text{Chamberlain}} - \boldsymbol{\beta}_{\text{RE\_Logit}}][\mathbf{V}(\boldsymbol{\beta}_{\text{Chamberlain}}) - \mathbf{V}(\boldsymbol{\beta}_{\text{RE\_Logit}})]^{-1}[\boldsymbol{\beta}_{\text{Chamberlain}} - \boldsymbol{\beta}_{\text{RE\_Logit}}]' \sim \chi^2_k \qquad [1]$$

where $\boldsymbol{\beta}_{\text{Chamberlain}}$ is the parameter vector corresponding to the covariates in Chamberlain's fixed effects logit model, $\boldsymbol{\beta}_{\text{RE\_Logit}}$ is the parameter vector for the comparable covariates in a random effects logit model, $\mathbf{V}(\cdot)$ denotes the relevant variance-covariance matrix for the two regression models, and k is the number of time-varying covariates in the estimated specifications. A statistically significant chi-squared test value implies the rejection of the random effects model in favour of the fixed effects alternative.

In order to determine the appropriateness of the Chamberlain estimator itself, we statistically test for the presence of the player-specific fixed effects. A Hausman-type test is also available for this purpose but constructed as the difference between Chamberlain's (conditional) ML estimates and the unconditional estimates from a conventional non-panel logit model. The estimates for the non-panel logit are obtained by pooling the data across all players, matches and seasons. Under the null hypothesis of no fixed effects, the ML estimates for this model are consistent and efficient. However, under the alternative, they are inconsistent since relevant determining factors are omitted. The estimates for Chamberlain's model are consistent under both the null and alternative hypotheses but inefficient under the former proposition. The relevant test statistic is expressed as follows:

$$[\boldsymbol{\beta}_{\text{Chamberlain}} - \boldsymbol{\beta}_{\text{Logit}}][\mathbf{V}(\boldsymbol{\beta}_{\text{Chamberlain}}) - \mathbf{V}(\boldsymbol{\beta}_{\text{Logit}})]^{-1}[\boldsymbol{\beta}_{\text{Chamberlain}} - \boldsymbol{\beta}_{\text{Logit}}]' \sim \chi^2_k \qquad [2]$$

where $\boldsymbol{\beta}_{\text{Chamberlain}}$ is as described above, $\boldsymbol{\beta}_{\text{Logit}}$ is the parameter vector for the comparable covariates from a pooled logit model, $\mathbf{V}(\cdot)$ denotes either the Chamberlain or the pooled logit variance-covariance matrix, and k is the number of time-varying covariates in the estimated specifications. If the individual fixed effects are relevant, then there will be a statistically significant difference between the Chamberlain ML estimates (that control for these effects) and the standard logit estimates (that do not). If the chi-squared test yields a statistically insignificant result, a standard (non-panel) pooled logit provides a valid alternative. A test for the presence of the random effects is more straightforward and requires use of a likelihood ratio test (LRT) where the unrestricted model is provided by the random effects logit. The restricted model is the standard non-panel logit with the data pooled across players, matches and playing seasons.

The effects of covariates influencing the award of a disciplinary card could potentially differ between ‘home’ and ‘away’ team players. Indeed, the empirical test for the presence of social pressure is based on statistically testing for such a difference. Although this concern could be accommodated and empirically evaluated using variables interacted with the status of the fixture (i.e., whether it is a ‘home’ or an ‘away’ fixture for the player), the player-specific effects included in such interacted models would still only reflect average player behaviour across both match contexts. This may not be entirely reasonable. For instance, ‘away’ team players may behave more aggressively than ‘home’ team players, thus justifying a higher caution rate and implying that the player-specific effects are not immutable across fixture type. In order to empirically investigate this issue, we compute a Chow version of the Likelihood Ratio Test (LRT) defined as:

$$\text{Chow LRT} = -2\times[L^{\text{Restricted}} - L^{\text{Unrestricted}}] \sim \chi^2_g \qquad [3]$$

where $L^{\text{Restricted}}$ is the maximized log-likelihood value based on the regression model (either fixed or random) pooled across fixture type, and $L^{\text{Unrestricted}} = L^{\text{‘home’}} + L^{\text{‘away’}}$ (i.e., the sum of the maximized log-likelihood values using data separated for the ‘home’ and ‘away’ team players respectively). The degrees of freedom $g = k^{\text{‘Home’}} + k^{\text{‘Away’}} - k^{\text{Restricted}}$ where k^j represents the number of parameters estimated for each of the j models. A significant chi-squared value for the test implies the rejection of a model pooled across ‘home’ and ‘away’ players and requires the estimation of models separated by the fixture status of the players’ teams.

Finally, matrix versions of Wald tests are used to test for the statistical significance of the joint effects for the sets of time–varying covariates relating to teams, referees and playing seasons in all the estimated models. The statistical significance of the time invariant factors of field position, region of origin and race included in the random effects model are also tested using the Wald testing principle.

Empirical Results

The first issue to be examined is whether the data empirically support the delineation of the player-level analysis across the 'home' and 'away' classification. This is assessed using the Chow variant of the LRT described in expression [3]. The final pair of prob-values reported in table 2 reveals that such a separation is justified for both models[7] and thus two sets of estimates are subsequently reported for each estimation procedure.[8] Although statistical evidence for the presence of player-specific random effects is detected using both the 'home' and 'away' data, the random effects model is rejected for both cases using the Hausman test outlined in expression [2]. However, the findings regarding the appropriateness of Chamberlain's fixed effects logit model are more nuanced given that, using expression [1], the implicit set of fixed effects in the 'home' specification is found to be statistically insignificant implying that a pooled (non-panel) logit model provides a more apposite specification for the 'home' team data. The Wald tests for all models reveal that referee and season effects are jointly significant at conventional levels, but that a team's playing style is not statistically important in determining the award of a disciplinary card. Statistically significant test values are also found for the random effects models for the time invariant factors corresponding to player field position in both 'home' and 'away' specifications, and for race and player region of origin in the former case only.[9]

The coefficients for the player and match-specific time-varying covariates are now the subject of discussion. The point estimate for the number of games played in the season to date is negative in all cases but found to achieve marginal statistical significance for only the

Chamberlain model fitted to the 'home' data. Given the corresponding estimate for this variable obtained using the pooled non-panel logit model (not reported here) is also found to be statistically insignificant, we conclude that the number of games played does not exert an independent effect on referee sanctioning behaviour. The estimated impact effects for those who remain on the field of play the entire duration of the match are broadly comparable for 'home' and 'away' team players. A player who plays the full 90 (or more) minutes of a match, and thus has a relatively higher risk exposure with respect to a disciplinary sanction, is approximately five percentage points more likely to receive a yellow card than a player who is either substituted or enters the game as a substitute.

Match officials are ostensibly more card-happy when the fixture is a derby match. On average and *ceteris paribus*, a player is about two percentage points more likely to be formally disciplined while playing in this type of fixture than in a non-derby game. The derby match estimates are broadly comparable across the two panel estimators and the 'home' and 'away' divide.[10] Although the estimated coefficient corresponding to the 'veteran' player variable is only found to be statistically significant in the random effects model using the 'home' player data, its sign and magnitude is in comport with the non-panel logit estimate obtained for this variable using the same data and specification. The significant estimate suggests that older players in the 'home' team are about 1.5 percentage points more likely to receive a yellow card than their younger counterparts, on average and *ceteris paribus*. The fact that players at the veteran stage of their careers are more likely to be sanctioned at 'home' than 'away' may be related to the fact that such players feel the need, more than younger players, to persuade home supporters and team management of their commitment to the club. This may be motivated by concerns over contract renewal/extension, and may induce such players to engage in more reckless challenges that invite a carding intervention.

The potential competitiveness of the fixture, as reflected in the relative league position of the two teams prior to the game, appears to matter more for the 'home' than the 'away' side regardless of the regression model used. In particular, the greater the positional advantage the home team enjoys in the league over the visitors prior to the fixture, the higher its carding probability. Again, the non-panel logit model yields a similar pattern of estimates to the panel models for this particular variable. It could be conjectured that the finding is attributable to the fact that lower ranked teams in the EPL, when playing a much higher ranked one away from home, adopt overly defensive and less open tactics in order to render the team more difficult to breach or break down. This may lead to frustration for the 'home' side players eventuating in the higher *ceteris paribus* carding rate observed.

Table 2 about here

The social pressure effects are now the subject of discussion. The proposition examined is whether crowd attendance, the proxy measure for social pressure, exerts a differential effect on referee carding behaviour depending on whether the player features in the 'home' or the 'away' team. The conjecture is that, if social pressure is present, stadium attendance should register no independent effect on the probability that a home team player receives a yellow card[11], but should exert a well determined positive effect on the carding of an 'away' team player. The key statistical test of interest relating to social pressure is interpretable as a right-sided one-tailed test for the 'away' team, with the test for the 'home' side treated as two-tailed in this case.

Table 3 summarises the relevant social pressure effects by replicating estimates reported in table 2 for the log attendance coefficients using the two panel models. The table now also includes the corresponding estimates for a non-panel pooled logit model. The prob-values for the statistical tests of interest are reported in the square brackets beneath the coefficients. The estimates selected to inform the empirical test for the presence of social pressure are derived from the econometric models that satisfied the relevant statistical tests reported in table 2. This array of tests revealed that the Chamberlin model was the most econometrically appropriate for the 'away' specification but that a non-panel pooled logit provided an adequate fit for the 'home' specification. This choice of estimates reveals that, on average and *ceteris paribus*, the carding rate of home players is found to be unrelated to the size of attendance. In contrast, the disciplining rate for 'away' team players is found to be positively related to attendance with a prob-value of 0.043 for the one-sided test. Thus, our preferred set of empirical estimates provides evidence congruent with the presence of a social pressure effect on EPL refereeing behaviour in the use of the commonest disciplinary sanction in professional soccer.

Table 3 about here

It is important to leaven this conclusion with a number of caveats. First, the absence of concordance in the key estimates across the different regression models that were fitted to the 'home' data prohibits an unambiguous inference that is robust across all the estimators used. This is a less clean and satisfactory state of affairs than we would wish. Nevertheless, we would reiterate that our choice of estimates is entirely governed by what the data and the econometric testing dictate. Second, and related to the above point, the conclusion offered on the presence of social pressure is contingent on being fully confident that the array of test statistics deployed to evaluate the different econometric models is adequate for their dedicated tasks. We believe the relatively large sample sizes used in the analysis provide some comfort that the required asymptotic conditions are satisfied in this regard. Third, despite the fact that two of the three 'home' estimates for attendance are found to be statistically insignificant, their numerical values dominate the corresponding point estimates for the three statistically significant 'away' effects.[12] This is concerning but highlights the negligible magnitude of the social pressure effects actually detected in this study. For instance, our preferred social pressure estimate suggests that an increase in attendance of 5%, assuming such a rise is actually feasible given stadium capacity, would increase the disciplinary rate by a modest 0.05 of one percentage point, on average and *ceteris paribus*. This modest change in attendance would induce a mere 0.45% increase in the yellow card rate for ‘away’ team players at the average, an effect which is unlikely to exert any meaningful influence on match outcomes.[13]

Conclusions

This paper uses player/match level data drawn from five seasons of the EPL to test for a refereeing susceptibility to social pressure in the application of the game’s commonest disciplinary sanction, the yellow card. The existing literature, using a diverse array of variables designed to capture refereeing bias, has generally reported evidence in comport with the presence of such an effect. However, we take the view that some of the studies in this literature have been characterised by deficiencies. For instance, we believe that existing research has used dubious measures to proxy for referee bias (e.g., ‘added time’ in close matches), questionable units of observation for the empirical analysis (e.g., match-level rather than player-level data), and inappropriate econometric techniques (i.e., OLS when the dependent variable outcomes are either truncated or limited in some way). It is unclear the extent to which the empirical findings on the presence of social pressure effects are sensitive to the these modelling issues, though the literature’s failure to engage with the above concerns clearly raises questions as to the veracity of existing findings on this theme.

Our focus on disciplinary sanction within the EPL follows on from the earlier work of Dawson *et al*. (2007), Buraimo *et al*. (2010) and Reilly and Witt (2011), but in its specific emphasis on social pressure is closer in spirit to the work of Dawson and Dobson (2010), Pettersson-Lidborn and Priks (2010), and Reilly and Witt (2013). We contend that in empirically modelling the formal disciplining of players by referees, the appropriate unit of observation is the player/match level. Such a focus permits the explicit treatment of unobservable player heterogeneity likely to impact the award of a sanction on the field of play, and also allows the introduction of game-specific factors and team-specific attributes. In addition, we further argue that adequate care is required in selecting the most appropriate econometric techniques given the nature of the data generally offered for use as the dependent variable in such applications. We believe that the current paper offers some value-added in regard to both of these concerns.

The central research question investigated in this paper asks whether referees in the EPL are susceptible to crowd coercion in disciplining ‘away’ team players. Using both player-specific fixed and random effects models and a non-panel pooled logit model, we uncovered some evidence for this proposition. Thus, the sanctioning behaviour of EPL referees over the time period reviewed here was found to be responsive to the exercise of social pressure mediated (presumably) through the crowd noise level. However, the estimated effect is found to be small in magnitude and unlikely to influence match outcomes in any meaningful way. On balance, our findings are best interpreted as providing suggestive evidence for the presence of a negligibly sized social pressure effect in one, albeit globally popular, top-tier European professional soccer league.

Footnotes

1. The size of the crowd attending a match is taken to capture the noise volume in the stadium. Nevill *et al*. (2002) present experimental evidence that some refereeing decisions in the English Premiership League (EPL) are influenced by crowd noise.

2. The natural experiment was provided by the fact that the Italian government forced clubs with deficient safety standards in both the Serie A and Serie B leagues to play their games in empty stadia. Only 21 of the 842 matches in their dataset were played in an empty stadium raising questions as to the integrity of the natural experiment. Furthermore, as noted by Buraimo *et al*. (2010), the government ruling tended to apply only to those clubs with a record of crowd trouble, so the outcomes examined may simply reflect club-specific rather than social pressure effects.

3. However, Mechtel *et al*. (2011), using data from the German Bundesliga for nine recent playing seasons, offer a more optimistic set of findings for sanction-depleted visiting teams.

4. Professional Game Match Officials Limited (PGMOL) was formed in 2001 to provide match officials for all professional football matches played in England. It is responsible for the training, development, assessment and monitoring of all referees in the professional game including the sub-set of professional referees who officiate at English Premiership League games. A professional referee officiating in the English Premiership earned about £49,000 per annum plus a match fee over the period covered by our analysis.

5. Buraimo *et al*. (2010), who use minute-level match data, provide a notable exception here.

6. Baltagi (2008, pp.237–244) provides an accessible account of Chamberlain's estimation procedure, and also discusses random effects versions of the panel models used here.

7. A significant Chow LRT value is also obtained for the non-panel pooled logit model with a prob-value of 0.002.

8. The estimated coefficients for both models are expressed in log odds ratio effects but can be roughly translated into probability effects through scaling by 0.087 (0.112) for the 'home' ('away') specifications.

9. See Reilly and Witt (2011) for a detailed discussion of the time-invariant factors, particularly those associated with race.

10. The estimates for both the match duration and derby game variables obtained using a non-panel pooled logit for the 'home' specifications are dimensionally comparable to those reported for a Chamberlain model fitted to these data.

11. The effect of social pressure, as measured by attendance, could conceivably exert a negative effect on the 'home' team carding rate though this is viewed as a considerably less plausible outcome.

12. Given the fact that there is overlap in the two sets of data points, t-tests cannot be used to test for the presence of statistical differences across the 'home' and 'away' divide used here.

13. The marginal effect is computed using the Chamberlain model estimate for the 'away' sample as 0.1032×0.112 (see footnote 8) and, given use of a log attendance measure, a 5% increase represents a change of 0.05. The overall effect is calculated as $0.1032 \times 0.112 \times 0.05 = 0.000578$. Therefore, a 5% increase in attendance induces a modest increase in the carding rate of about 0.06 of one percentage point. This number should be interpreted relative to the average carding rate of approximately 0.13 (see table 1) and thus represents a 0.4% increase in the carding rate relative to this average.

References

Akerlof, G.A. (1980) A Theory of Social Custom of Which Unemployment May be One Outcome, *Quarterly Journal of Economics*, 94, 749-775.

Baltagi, B.H. (2008) *Econometric Analysis of Panel Data*, 4th Edition, Wiley.

Becker, G.S. and K.M. Murphy (2000) *Social Economics: Market Behaviour in a Social Environment* (Cambridge, MA: Harvard University Press).

Bernheim, D.B. (1994) A Theory of Conformity, *Journal of Political Economy*, 102, 841-877.

Boeri, T. and B.Severgnini (2011) Match Rigging and the Career Concerns of Referees, *Labour Economics*, 18(3), 349-359.

Boyko, R.H., Boyko, A.R., and M.G.Boyco (2007) Referee Bias Contributes to Home Advantage in English Premiership Football, *Journal of Sports Science*, 25, 1185-1194.

Buraimo, B., Forrest, D., and R.Simmons (2010), The 12th man? Refereeing Bias in English and German Soccer, *Journal of the Royal Statistical Society (Series A)*, 173 (part 2), 431 – 449.

Chamberlain G. (1980) Analysis of Covariance with Qualitative Data, *Review of Economic Studies*, XLVII, 225-238.

Dawson, P. and S. Dobson (2010) The Influence of Social Pressure and Nationality on Individual Decisions: Evidence from the Behaviour of Referees, *Journal of Economic Psychology*, 31, 181-191.

Dawson, P., Dobson, S., Goddard, J. and J. Wilson (2007) Are Football Referees Really Biased and Inconsistent? Evidence on the Incidence of Disciplinary Sanction in the English Premier League, *Journal of the Royal Statistical Society, Series A*, 170, 231-250.

Dohmen, T.J. (2008) The Influence of Social Forces: Evidence from the Behaviour of Football Referees, *Economic Inquiry*, 46, 411-424.

Garicano, L., Palacios-Huerta, I., and C. Prendergast (2005) Favouritism under Social Pressure, *Review of Economics and Statistics*, 87, 208-216.

Mechtel, M., Bäker, A., Brändle,T., and K.Vetter (2011) Red Cards: Not Such Bad News for Penalized Guest Teams, *Journal of Sports Economics*, 12, 621-646.

Nevil, A.M., Balmer, N.J., and A.M.Williams (2002) The Influence of Crowd Noise and Experience upon Refereeing Decisions in Football, *Psychology of Sport and Exercise*, 3, 261–272.

Pettersson-Lidbom, P., and M.Priks (2010) Behavior under Social pressure: Empty Italian Stadiums and Referee Bias, *Economics Letters*, 108, 212-214.

Reilly, B., and R. Witt (2011) Disciplinary sanctions in English Premiership football: Is there a racial dimension? *Labour Economics*, 18, 360-370.

Reilly, B., and R. Witt (2013) Red cards, referee home bias and social pressure: evidence from English Premiership soccer, *Applied Economics Letters*, 2013, Vol. 20, No.7, pp.710-714.

Rickman, N. and R. Witt (2008) Favouritism and Financial Incentives: A Natural Experiment, *Economica*, 75, 296-309.

Ridder, G., Cramer, J.S., and P. Hopstaken (1994) Down to Ten: Estimating the Effect of a Red Card in Soccer, *Journal of the American Statistical Association*, 89, 1124-1127.

Table 1: Variable Descriptions and Summary Statistics

Variable Name	Variable Description	All	'Home'	'Away'
Yellow Card	= 1 if the player received a yellow card in a match; = 0 otherwise.	0.1124	0.0959	0.1289
Prior Games Played	The number of games played by the player in the current season prior to the current game.	13.464 (9.129)	13.527 (9.145)	13.402 (9.114)
Played 90 Minutes	= 1 if the player played for the entire game; = 0 otherwise.	0.6317	0.6316	0.6318
'Veteran' Player	= 1 if the player is over 32 years of age at the start of the relevant season; = 0 otherwise.	0.1173	0.1186	0.1160
Relative League Position	The difference in league position between the home team and the away team prior to the start of the current game.	0.2026 (5.779)	0.1651 (8.260)	0.2401 (8.261)
'Derby'	= 1 if the game is a 'derby' game; = 0 otherwise.	0.1347	0.1347	0.1347
Home	= 1 if the game is a home game; = 0 otherwise.	0.5003	1.0	0.0
Ln(Attendance)	The natural log of the attendance at the game.	10.390 (0.354)	10.391 (0.353)	10.390 (0.354)
N	The Number of Observations	51,076	25,552	25,524

Notes to table 1: (a) The total sample of 51,076 match-level observations on 1162 players over five seasons is used in the calculation of the summary statistics. (b) Standard deviations are reported in parentheses for the continuous variables only. (c) The 30 premiership clubs for the five seasons are: Arsenal, Aston Villa, Birmingham City, Blackburn Rovers, Bolton, Charlton Athletic, Crystal Palace, Chelsea, Derby County, Everton, Fulham, Liverpool, Leeds United, Leicester, Manchester United, Manchester City, Middlesboro, Newcastle United, Norwich City, Portsmouth, Reading, Sheffield United, Southampton, Sunderland, Tottenham Hotspur, Watford, West Bromwich Albion, West Ham United, Wigan Athletic, and Wolverhampton Wanderers. (d) The 'Home' column is for observations for the 'home' side and the 'away' column is for observations for the visiting side.

Table 2: The Determinants of Receiving a Disciplinary Sanction (i.e., a Yellow Card)

Variables	Fixed Effects Logit		Random Effects Logit	
	'Home'	'Away'	'Home'	'Away'
Constant	†	†	-5.4492** (2.5447)	-3.4929*** (0.6970)
Prior Games Played	-0.0044* (0.0026)	-0.0027 (0.0023)	-0.0037 (0.0025)	-0.0029 (0.0022)
Played 90 Minutes	0.5505*** (0.0567)	0.4885*** (0.0492)	0.6010*** (0.0532)	0.5374*** (0.0470)
'Veteran' Player	-0.0488 (0.1553)	0.0389 (0.1370)	0.1818** (0.0891)	0.0041 (0.0854)
Relative League Position	0.0199*** (0.0033)	-0.0027 (0.0031)	0.0189*** (0.0032)	-0.0034 (0.0031)
'Derby' Game	0.2034*** (0.0644)	0.2275*** (0.0561)	0.1924*** (0.0637)	0.2390*** (0.0557)
Ln(Attendance)	0.5463** (0.2556)	0.1032* (0.0603)	0.3109 (0.2356)	0.1101* (0.0598)
Prob-values for Statistical Tests:				
Prob-value for Team Effects	0.1572	0.1518	0.2864	0.4144
Prob-value for Referee Effects	0.0000	0.0000	0.0000	0.0000
Prob-value for Season Effects	0.0059	0.0022	0.0085	0.0132
Prob-value for Field Positions	†	†	0.0000	0.0000
Prob-value for Race	†	†	0.0132	0.3939
Prob-value for Region of Origin	†	†	0.0000	0.6674
Prob-value for Fixed Effects	0.1289	0.0166	†	†
Prob-value for Random Effects	†	†	0.0000	0.0000
Prob-value for Hausman Test	†	†	0.0031	0.0083
Prob-value for Chow LR Test	0.0000		0.0015	
Number of Players	662	737	1,104	1,125
Observations	21,371	22,672	25,552	25,524
Log-Likelihood Value	-5949.8	-7436.0	-7609.4	-9297.6

Notes to table 2: (a) ***, ** and * denote statistical significance at the 0.01, 0.05 and 0.10 levels respectively using two-tailed tests. (b) † denotes not applicable in estimation. (c) In the Chamberlain fixed effects logit model 442 (388) players comprising 4,181 (2852) observations were dropped because of either all positive or all negative outcomes for the case of 'home' ('away') fixtures. The log-likelihood values for the two logit models are not directly comparable. (d) The random effects models also contain time invariant factors capturing a player's field position, racial group, region of origin, and whether English is a player's native language. (e) The Prob-values for the Team, Referee and Season Effects are based on the matrix forms of Wald tests for the overall significance of these effects in the relevant regression models. (f) The test for the presence of fixed effects is based on expression [1] in the text. The Hausman test is based on expression [2] in the text. The Chow LR Test is based on expression [3] in the text.

Table 3: The Estimated Social Pressure Effects

	Pooled Logit	**Fixed Effects Logit**	**Random Effects Logit**
'Home'	0.1838	0.5463	0.3109
Prob-values for Tests	[0.397]	[0.033]	[0.187]
'Away'	0.1095	0.1032	0.1101
Prob-values for Tests	[0.031]	[0.043]	[0.033]

Notes to table 3: (a) The log attendance estimates for the fixed and random effects logit models are based on the regression models reported in table 2. The pooled logit estimates are based on comparable specifications to the models in table 2 but do not explicitly control for omitted player effects. The full results for these pooled logit models are not reported here for brevity. (b) The prob-values for the 'home' specifications are based on two-tailed tests, while the prob-values for the 'away' specifications are based on one-tailed tests (see text).