IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17512

Inequity Aversion in Subjective Evaluations: Evidence from Referees' Decisions in Soccer

Leo Morabito
Vincenzo Scoppa

DECEMBER 2024

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17512

Inequity Aversion in Subjective Evaluations: Evidence from Referees' Decisions in Soccer

Leo Morabito
University of Milan

Vincenzo Scoppa
University of Calabria and IZA

DECEMBER 2024

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9 53113 Bonn, Germany	Phone: +49-228-3894-0 Email: publications@iza.org	www.iza.org

ABSTRACT

Inequity Aversion in Subjective Evaluations: Evidence from Referees' Decisions in Soccer*

Subjective evaluations in many contexts might be affected by decision-makers' social preferences. To explore this phenomenon, we use data from soccer referees' decisions. According to soccer rules, referees are expected to evaluate each episode independently, without taking into account previous decisions. However, if referees are averse to creating inequities between teams, they might seek to balance their decisions and, for example, after awarding a penalty to the home team, they may raise the evidence threshold for awarding a second penalty to the same team, while lowering it for awarding a penalty to the away team. First, we offer a simple theoretical model to explain these insights. Then, using detailed minute-by-minute commentary data from approximately 21,400 matches in major European leagues, we show a strong preference by referees to treat teams fairly: they reduce the probability of awarding a penalty (or a red or yellow card) to a team if it has already been awarded to that team, while increasing the probability if it has been awarded to the opposing team. In the final part, focusing on injury time, we show that referees tend to lengthen injury time both when the home team is behind and when the away team is behind, suggesting that referees may have a preference for treating both teams fairly.

JEL Classification: D91, L83, Z20, D63, Z28

Keywords: subjective evaluations, social preferences, inequity aversion, compensatory behavior, social pressure, soccer, behavioral economics

Corresponding author:
Vincenzo Scoppa
Department of Economics, Statistics and Finance
University of Calabria
Via Ponte Bucci
87036 Arcavacata di Rende (CS)
Italy
E-mail: v.scoppa@unical.it

* We would like to thank Carlo Devillanova, Marco Alberto De Benedetto, Maria De Paola, Davide Fiaschi, Balint Menyhert, Giuseppe Migali, Paolo Naticchioni, Roberto Nisticò, Francesco Principe and participants of the online ROSES seminar, particularly James Reade, Peter Dawson, Stefan Szymanski, Alex Krumer, Raphael Flepp, Jorge Tovar, as well as seminar participants at the 2024 AIEL Conference (Naples), for their useful comments and suggestions.

1. Introduction

Subjective evaluations are widespread in many contexts, such as when evaluating candidates for hiring or employees for a promotion in firms or organizations, in assessing students' proficiency in schools or universities, in judging people accused of crimes, and so on. These kinds of evaluations might be affected by cognitive biases (such as availability or representativeness biases) or by the decision-maker's preferences, such as the desire to favor one subject instead of another (favoritism), the preference for treating parties fairly, the desire for social approval.

A large body of literature in behavioral economics has shown that individuals are not only concerned with their own material payoffs, but they are also typically averse to inequality and strive to maintain a positive social image (see, among others, Charness and Rabin, 2002; Fehr and Schmidt, 1999; Bénabou and Tirole, 2006). Inequity aversion models show that individuals often avoid outcomes perceived as unfair, even at their own expense, while models of social pressure show that decision-makers tend to behave according to what others desire or consider appropriate.

In real world contexts, it is challenging to analyze whether and how evaluations are affected by equity concerns or social pressure, since the merit of each decision, the decision-maker's available information, the pressure they face, are typically unobservable.

Thanks to the availability and richness of data, sport contests are a fruitful field for empirically testing these aspects. For example, during a soccer match, referees take a large number of decisions under varying contexts and circumstances, offering a unique setting to examine individual biases and preferences in action.

One widely studied factor in sports is social pressure, the tendency of referees to conform to the pressure of the crowd, most of the time supporting the home team. In this paper we focus on a relatively less explored aspect: the desire of decision-makers to be fair to all the parties, or the desire to be perceived as fair.

According to the rules of soccer, referees should evaluate each situation in a match independently, without taking into account previous decisions, the current score, the identity of the teams, etc. However, referees' decisions may be influenced by their aspiration to maintain impartiality. If referees have a preference to treat teams fairly or to be perceived as fair by players, spectators, experts and committees evaluating their performance, or are averse to influencing the final score with a wrong call, they might try to balance their decisions throughout the match. They may consciously or unconsciously seek to equalize prior decisions made in favor of or against a team. A vivid example of this attitude is reported by "*Corriere dello Sport*", an Italian sports newspaper: "At the end of the "Serie A" match Juventus-Bologna on August 27, 2023, Bologna's team manager, Marco Di Vaio asked the referee Marco Di Bello "Why didn't you award that penalty to Bologna?". The referee replied: "Because I previously didn't award a penalty to Juventus for a foul on Chiesa [Juventus player]".

A related explanation, suggested by Chen *et al.* [2016], is that referees – in their desire to be fair to both parties – might prefer to alternate being "mean" and "nice" during a match. If the referee is affected by the "law of small numbers" – thinking that a short sequence is representative of the whole process – he could over-infer, for example, from a short sequence of mean decisions that he is becoming too negative towards a team and decide to reverse his behavior.

Using data from five major European soccer leagues (England, Spain, Germany, France, Italy) observed for 12 seasons (from 2012/13 to 2023/24), totaling about 21,400 matches, we analyze the behavior of referees, focusing on minute-by-minute commentary (for a total of about 2.1 million observations). In our main analysis we analyze the probability that in each time frame the referee takes disciplinary sanctions (penalties, red and yellow cards) in favor of a team or of its opponent, on the basis of his previous decisions of the same type. We control a number of measures of defensive and attacking tactics adopted by the two teams in each phase of the match (current score, recent and cumulated shots, fouls, and so on) and for the quality of the two competing teams (using the probability of winning based on betting odds or teams' Elo ratings).

We find that the probability of awarding an additional penalty in favor of a team is reduced if previously the same team has been given a penalty, while the probability is increased if a penalty has been awarded to the opposing team. This holds true for decisions favoring the home team but also the away team. A similar compensatory pattern emerges with respect to red and yellow cards.

In the final part of the paper, we consider the referees' decisions on injury time (or extra time) allowed at the end of the match, which should depend only on time lost during the match for substitutions, injuries, protests, and so on. Several papers have found that referees succumb to social pressure by allowing additional injury time when the home team is behind. We confirm this finding, but in addition we find that referees extend injury time also when the away team is trailing. This is in stark contrast with the simple explanation based on crowd pressure, but it is in line with the explanation that referees aim to be impartial and to give a chance to equalize to whichever team is losing.

The paper is organized as follows. In Section 2, we briefly discuss the related literature. In Section 3 we propose a very simple theoretical model of referees' decisions to explain their behavior. Section 4 describes the data. In Section 5, we carry out our econometric analysis of minute-by-minute decisions of referees, regarding penalties, red cards and yellow cards and run some robustness checks. In Section 6 we study whether the uncovered effects are heterogeneous under different contexts. Section 7 examines referees' decision on injury time on the basis of the score at the 90th minute. Section 8 concludes.

2. Related Literature

Most analyses in sports on referees' decisions focus on the effects of social pressure, arguing that referees can be subconsciously influenced by the noise of a large crowd in the stadium and react by favoring the home team, awarding it more penalties, fewer disciplinary sanctions, and more injury time when it is behind (for an exhaustive survey, see Dohmen and Sauermann, 2016).

Dawson *et al.* (2007), among others, document that referees show favoritism towards the home team in awarding yellow and red cards in the Premier League. Additionally, Dawson and Dobson (2010) find that social pressure (as well as nationality) affects discretionary decisions by referees in European cup matches.[1] Reade, Schreyer and Singleton (2022) and Scoppa (2021), among others, exploiting the natural experiment with empty stadiums following the COVID-19 pandemic, show that referees without the pressure of the crowd are much more balanced in awarding penalties, yellow and red cards to home and away teams.

Garicano, Palacios-Huerta and Prendergast (2005) show that the amount of injury time that referees assign at the end of a match is roughly twice as much when it is likely to advantage the home team (since it is one goal behind) than when it is likely to hurt it (when it is one goal ahead). A similar result has been found by Dohmen (2008) and Sutter and Kocher (2004) for the German Bundesliga, and by Scoppa (2008) for the Italian "Serie A".

More related to the present study, some works have analyzed whether referees' decisions are influenced by their aspiration to maintain impartiality in the match and by their aversion to inequity. Plessner and Betsch (2001) examine sequential effects in referees' decisions regarding penalty kicks in soccer, focusing on whether referees' decisions are influenced by previous calls within the same match. They conducted a lab experiment asking their sample of participants, composed of professional referees and soccer players, to act as referees and decide on 20 videotaped scenes from a soccer match, some of which contained ambiguous foul situations inside the penalty areas of both teams. The researchers manipulated the order of the ambiguous scenes, to create different conditions. They found that referees were less likely to award a penalty to the same team consecutively. Specifically, if a penalty had already been awarded to one team, then that team was less likely to be awarded another penalty, while the other team was more likely to be awarded one in a subsequent ambiguous situation. As in many lab experiments, these findings might be influenced by the relatively pressure-free context (Mascarenhas *et al.*, 2002). Moreover, video scenes are conditioned by the spectator's – or sideline – viewing position, which differs from the typical perspective taken by a referee.

Schwarz (2011) investigates the existence of a compensatory tendency in soccer referees' decisions, using a dataset of matches from the German Bundesliga, with a focus on 441 matches that involved exactly two penalty kicks. The study compares the observed distribution of penalties to the expected distribution under the assumption of independent penalty decisions and finds strong evidence that matches in which each team is awarded one penalty are more likely than expected under independent referee decisions. This suggests that referees are likely to balance their decisions to avoid being perceived as biased. Schwarz (2011) also argues that the players on the field can perceive a clear error by the referee and try to exploit it by exaggerating or simulating a fall in the penalty area or simply appealing to the referee's impartiality, inducing the referee to make an erroneous call that evens the score.[2]

---

[1] In a lab experiment Nevill *et al.* (2002) compared professional referees' decisions taken watching a videotaped recording of a match when they hear the reactions of the crowd with their decisions taken when they watch the silent video. They found that referees exposed to the crowd noise were more acquiescent to the home team and in line with the effective decisions taken on the field. Referees have to make instantaneous decisions, and they tend to focus on the most salient cues, one of which may be the crowd noise. Moreover, they tend to avoid potential displeasure to the crowd.

[2] Chen *et al.* (2016) examined the tendency for sequential decisions to exhibit negative autocorrelation pattern in high-stakes contexts, such as asylum judgments, loan approvals and baseball umpire calls. They found that judges were

Along the same lines, Moskowitz and Wertheim (2011), in analyzing Major League Baseball pitches suggest that "when an obviously bad call is made, the officials soon compensate it by making an equally bad call that favors the other team. Or, in the next ambiguous situation, the referee will side with the team that was wronged previously".

Anderson and Pierce (2009) investigate whether referees in National Collegiate Athletic Association basketball exhibit a tendency to even out the number of fouls between competing teams during a game. The study uses data from 365 basketball games during the 2004-2005 season. The authors apply a logistic regression model to analyze the probability of foul calls, considering factors such as the net difference in fouls between teams, the current score, and whether the game is played at a neutral venue. Their analysis reveals a significant tendency for referees to call fouls against the team with fewer fouls committed. A similar analysis conducted by Noecker and Roback (2012), using data from both 2004-2005 and 2009-10 seasons, confirms their results.[3]

More recently, Considine *et al.* (2024) contributed to this literature using data from 75 hurling games played in Ireland between 2016 and 2018. The authors employ Probit models to examine whether free shots awarded by referees are determined by the score margin or the net free count at the time. They find clear evidence of a compensatory tendency: referees are more likely to award a free shot to the team that is behind in the free count (or the score).

With respect to this body of literature, we use a much larger dataset (containing over 2 million minute-by-minute observations), consider several referee decisions (penalties, red cards, yellow cards, injury time), and are able to estimate the probability of a referee's decision on the basis of his prior decisions, accounting for a range of factors that could influence it – such as the tactics adopted by teams in each phase of the match, the quality of the teams, referee fixed effects, etc.

3. A Simple Model of Referee Decisions with Inequity Aversion

In this Section, we propose a simple model to help in interpreting referees' decisions. We assume a referee's utility function with an aversion to making mistakes – mainly because of career concerns (since evaluators of their performance might penalize them) or to avoid disapproval from players, spectators, experts, etc. – and, inspired by the models of inequity aversion proposed by Charness and Rabin (2002) and Bolton and Ockenfels (2000), we assume that referees also exhibit an aversion to treating teams unequally:

$$U = -Disutility\ from\ mistakes - \theta(Imbalance\ created)^{\gamma}$$

---

significantly more likely to deny asylum if they granted asylum to the previous applicant, with this effect being stronger following longer streaks of similar decisions. Similarly, loan officers are more likely to deny a loan application after approving the previous one. They interpret these findings as a consequence of the law of small numbers and gambler's fallacy that would affect decision-makers.

[3] Haynes and Gilovich (2010) found that even players might show a compensatory bias. They found that basketball players tend to miss more free throws when they are the result of incorrect calls by the referee, indicating their conflicting behavior in response to perceived inequities.

θ represents the degree of aversion to creating an imbalance ($\theta \geq 0$) or aversion to influencing the score. When $\theta = 0$ referees are not concerned about inequity and focus solely on avoiding mistakes. We assume $\gamma > 1$ to represent increasing sensitivity to inequity with a convex function.

Suppose an episode during the match gives rise to a penalty (or a red or a yellow card) for Team 1 with probability p. Let's evaluate the referee's utility when he awards and when he denies the penalty. The utility of awarding the penalty is:

$$U_A = p(0) + (1 - p)(-\phi) - \theta$$

With probability p the decision is correct and the referee incurs no disutility, while with probability $(1 - p)$ the decision is wrong (a Type I error or false positive) resulting in a disutility of ϕ. Additionally, since the penalty creates an advantage for Team 1, the referee experiences an extra disutility of θ.

On the other hand, the utility when the referee does not award the penalty is:

$$U_N = p(-\phi) + (1 - p)0$$

With probability p the decision is wrong (a Type II error or false negative), resulting in a disutility of ϕ. However, by denying the penalty, the referee does not create an imbalance between the teams and incurs no further disutility.

The referee chooses to award the penalty if $U_A > U_N$, which gives:

$$p > \frac{\phi + \theta}{2\phi} = \tilde{p}$$

If referees were not concerned about inequity ($\theta = 0$), awarding a penalty would require only that $p > 0.5$. However, when referees are averse to treating teams unequally – or to being decisive in determining the final outcome of the match, $\theta > 0$ – the probability must exceed the threshold $\tilde{p}$, which is greater than 0.5.

Awarding a penalty to the opposing team

Suppose a penalty has already been awarded to Team 1. Now consider an episode during the match in favor of the opposing Team 2, which results in a penalty with probability q. The referee's utility for awarding the penalty is:

$$U_A = qp(0) + q(1 - p)(-\phi) + (1 - q)p(-\phi) + (1 - q)(1 - p)(-2\phi)$$

With probability qp, no mistakes are made; with probability $q(1 - p)$ and $(1 - q)p$ only one mistake occurs, while with probability $(1 - q)(1 - p)$ two mistakes are made. Since awarding the second penalty to Team 2 compensates for the first penalty awarded to Team 1, the disutility associated with treating teams unequally is eliminated.

The referee's utility for not awarding a penalty is as follows:

$$U_N = qp(-\phi) + q(1 - p)(-2\phi) + (1 - q)p(0) + (1 - q)(1 - p)(-\phi) - \theta$$

The referee awards the second penalty if $U_A > U_N$, that is:

$$q > \frac{\phi - \theta}{2\phi} = \tilde{q}$$

To award a penalty to the opposing team, following a penalty awarded to the first team, the evidence threshold is lower, $\tilde{q} < \tilde{p}$. Therefore, referees are more likely to award a penalty to a team if a penalty has previously been awarded to the opposing team.

Awarding a second penalty to the same team

Suppose a penalty has been awarded to Team 1. Now consider a second episode in favor of the same Team 1, which results in a penalty with probability *r*. The utility of awarding the penalty is:

$$U_A = rp(0) + r(1-p)(-\phi) + (1-r)p(-\phi) + (1-r)(1-p)(-2\phi) - \theta 2^{\gamma}$$

The disutility resulting from the creation of imbalance is greater, since two decisions are made against Team 2.

The utility for not awarding a penalty is:

$$U_N = rp(-\phi) + r(1-p)(-2\phi) + (1-r)p(0) + (1-r)(1-p)(-\phi) - \theta$$

The referee awards the second penalty to Team 1 if $U_A > U_N$, that is:

$$r > \frac{\phi + \theta(2^{\gamma} - 1)}{2\phi} = \tilde{r}$$

This result indicates that, to award a second penalty to Team 1, the evidence threshold must be higher. Consequently, referees are less likely to award a penalty to a team if a penalty has previously been awarded to the same team.

In summary, the empirical predictions of our model are:

$$\tilde{q} < \tilde{p} < \tilde{r}$$

This means that referees are more likely to award a penalty to a team if the opposing team has previously been awarded a penalty, whereas they are less likely to award a penalty to a team if the same team has already been awarded a penalty.

4. Data and Descriptive Statistics

The dataset we use includes 21,552 matches from the top five European soccer leagues (English "Premier League", Spanish "La Liga", German "Bundesliga", Italian "Serie A", French "Ligue 1") spanning 12 consecutive seasons, from 2012-13 to 2023-24. Each league is composed of 18-20 teams. In each season, teams play against each other twice (once at home and once away) resulting in a total of 34-38 matches.[4] Table A1 in the Appendix presents the number of observations by league and season.

Our analysis focuses on match commentary generated from live play, sourced from ESPN

[4] In the first half of the season, each team plays against all its opponents once, approximately alternating between home and away matches; in the second half, each team plays against the same teams, but home and away matches are reversed.

(www.espn.com).[5] From these commentaries, we extracted all relevant events for both the home team and the away team, along with the exact minute they occurred: goals, penalties, red cards, yellow cards, shots, substitutions, fouls, free kicks, corners, offsides, injuries, VAR (Video Assistant Referee) episodes. Additionally, for each match we observe the date, venue, injury time allowed in both halves, attendance and the identity of the referee.

We cross-checked our data with the data at match level collected from the website www.football-data.co.uk, and we removed 132 matches (representing 0.6% of the sample) that contained inconsistencies in goals, penalties, or red cards. Our final dataset consists of 21,420 matches.

In our analysis, we partition each match into 100 equal time frames, each approximately one minute long, since soccer matches often extend beyond 90 minutes due to injury time. This method follows the percentile-based segmentation proposed by Robberechts, Van Haaren, and Davis (2019), ensuring that each segment of the match is equally represented, with the end of the first half corresponding to the 50th percentile. This approach enables us to capture referee decisions minute by minute, as a function of his prior decisions and of the dynamic nature of the match – since teams adapt their strategies by adjusting their attacking and defensive tactics.

Table 1 reports the descriptive statistics using the data in which each match is divided into 100 time intervals. On average, home teams are awarded 0.18 penalties per match (or 0.0018 per time frame), while

Table 1. Descriptive Statistics: Minute-by-Minute Data

Variable	Mean	Std. Dev.	Min	Max	Obs.
Penalty Home	0.0018	0.0424	0	1	2142000
Penalty Away	0.0013	0.0357	0	1	2142000
Red Card Home	0.0009	0.0304	0	2	2142000
Red Card Away	0.0012	0.0341	0	2	2142000
Yellow Card Home	0.0196	0.1402	0	4	2142000
Yellow Card Away	0.0224	0.1496	0	4	2142000
Score Diff. Home-Away	0.1604	1.2027	-9	9	2120580
Cum. Penalties Home	0.0785	0.2813	0	3	2142000
Cum. Penalties Away	0.0555	0.2373	0	3	2142000
Cum. Red Cards Home	0.0276	0.1686	0	3	2142000
Cum. Red Cards Away	0.0354	0.1902	0	3	2142000
Cum. Yellow Cards Home	0.7543	1.0227	0	9	2142000
Cum. Yellow Cards Away	0.8749	1.1101	0	9	2142000
Cum. Fouls Home	5.8239	4.3306	0	32	2142000
Cum. Fouls Away	5.9788	4.4230	0	31	2142000
Cum. Shots Home	6.4808	5.1473	0	44	2142000
Cum. Shots Away	5.2305	4.3597	0	37	2142000
Cum. Video Reviews	0.0973	0.3488	0	7	2142000
Prob. Home Team Win	0.4682	0.1930	0.038	0.980	2141400
Elo Diff. Home-Away	0.2349	157.35	-532.81	540.4	2142000

Notes: Data from the First Divisions of England, Spain, Italy, Germany and France. Seasons: 2012-2013 to 2023-2024. Data collected from ESPN. Observations are organized at the 1-minute frame level.

[5] An example of commentary from the match Real Madrid-Barcelona (played on April 21, 2024, available at: https://www.espn.com/soccer/commentary/_/gameId/674330): “17’: Penalty Real Madrid. Lucas Vázquez draws a foul in the penalty area. Penalty conceded by Pau Cubarsí (Barcelona) after a foul in the penalty area; Goal! Real Madrid 1, Barcelona 1. Vinícius Júnior (Real Madrid) converts the penalty with a right footed shot to the bottom right corner”.

away teams receive 0.13 penalties (or 0.0013 per time frame). On average, home teams are issued 0.09 red cards and 1.96 yellow cards per match, while away teams receive 0.12 red cards and 2.24 yellow cards.

For each time frame t in a match, we calculate at time $t - 1$ the score and the total cumulative counts of penalties, red and yellow cards, fouls, shots, corners and offsides for both the home and away teams.

To take into account the quality of the teams, we consider two measures. First, we use the betting odds of an online gambling company, *Bet365*.[6] Since betting odds reflect the inverse of the probability of an outcome, we calculate the probability of winning for the home team as: $Prob. Home\ Team\ Win = (1/Betting\ Odds\ Home\ Team)$. We also use Elo ratings, which measure team strength based on past results, updated weekly from https://www.clubelo.com, and construct the variable *Elo Difference Home-Away*.

Furthermore, we collected match-level data on attendance and referee identity (age, experience, number of red and yellow cards, number of penalty kicks, etc.) from www.FBstat.com and www.transfermarkt.com.

Correlations of referee decisions for home and away teams

As preliminary evidence of referees' tendency to balance decisions between the two teams, we examine the correlation between penalties awarded to the home team and to the away team at the match level (21,420 observations). We run a simple regression of *Penalty Home* on *Penalty Away*, controlling for referee fixed effects to account for any difference in attitudes across referees (and leagues). Additionally, we include season fixed effects to take into account possible differences across seasons. If these decisions were taken by referees independently, we should observe no correlation. In column (1) of Table 2 we find that penalties awarded to the home and away teams are positively correlated (coefficient=0.032; *t*-stat=3.55).

We conduct a similar regression for red cards awarded to the home and away teams, and again we find a positive coefficient (+0.056), with a *t*-stat of 6.44 (column 2). In addition, in regressing *Yellow Cards* for the home and away teams, we find a coefficient of +0.153 and a *t*-stat of 21.2 (column 3).

Remarkably, the positive correlations between home and away teams in referees' decisions stand in stark contrast to the negative correlations observed for other match outcomes: in column (4) we show a strong negative correlation between goals scored by home and away teams (*t*-stat=–10.7), in column (5) a negative correlation between home and away shots (*t*-stat=–34.5), and in column (6) a negative correlation between home and away corners (*t*-stat=–35.5).

This evidence suggests that referees – when taking disciplinary decisions – tend to balance these decisions between the opposing teams. In contrast, key match outcomes such as goals and shots indicate that the attacking and defensive dynamics are not balanced between the two teams during a match.

[6] Betting odds represent the amount that a bettor would win relative to the stake, should he or she predict the correct outcome (e.g. home win, draw, away win).

Table 2. Correlations between Referees' Decisions for Home and Away Teams and Match Outcomes (Goals, Shots and Corners) at the Match Level

	(1) Penalties Home	(2) Red Cards Home	(3) Yellow Cards Home	(4) Goals Home	(5) Shots Home	(6) Corners Home
Penalties Away	0.032***					
	(0.008)					
Red Cards Away		0.056***				
		(0.009)				
Yellow Cards Away			0.153***			
			(0.007)			
Goals Away				-0.092***		
				(0.011)		
Shots Away					-0.378***	
					(0.014)	
Corners Away						-0.286***
						(0.008)
Referee F.E.	YES	YES	YES	YES	YES	YES
Observations	21420	21420	21420	21420	21420	21420
Adjusted R^2	0.006	0.014	0.108	0.012	0.132	0.079

Notes: The table reports OLS estimates. The dependent variables are listed at the top of each column. Standard errors (reported in parentheses) are adjusted for heteroskedasticity and clustered at the referee level. The symbol *** indicates that the coefficients are statistically significant at the 1% level.

5. An Econometric Analysis of Minute-by-Minute Referee Decisions

In this Section, we carry out a more comprehensive econometric analysis of referee decisions, using minute-by-minute events to assess whether referees base their decisions on the basis of previous decisions made in favor of or against the home and away teams. We control for the current score as well as for several variables that account for the quality of the teams and the dynamic strategies they are adopting.

The main equations we estimate, for both the home (*H*) team and away (*A*) teams, are as follows:

$$Y_{itH} = \beta_0 + \beta_1\left(\sum_{\tau=1}^{t-1} Y_{i\tau H}\right) + \beta_2\left(\sum_{\tau=1}^{t-1} Y_{i\tau A}\right) + \beta_3 ScoreDiff_{it-1} + \beta_4 Cum.\boldsymbol{X}_{it-1} + \beta_5 \boldsymbol{W}_i + \beta_6 \boldsymbol{Z}_i + u_{itH}$$

where Y_{itH} represents a binary decision by the referee in match *i* at minute *t* (for the home team) as regards to penalties, red cards and yellow cards. The term $\sum_{\tau=1}^{t-1} Y_{i\tau H}$ captures the cumulated decisions of the same type made prior to time t for the home team, while $\sum_{\tau=1}^{t-1} Y_{i\tau A}$ is the corresponding variable for the away team. $ScoreDiff_{it-1}$ represents the difference in score between the home and the away teams one minute before the event under consideration; $Cum.\boldsymbol{X}_{it-1}$ is a vector of variables measuring cumulative or recent events related to attacking or defensive tactics (such as number of shots, fouls, corners, offsides, etc.); $\boldsymbol{W}$ is a vector of variables measuring the quality of the opposing teams, and $\boldsymbol{Z}$ includes other controls, such as referee fixed effects, the availability of VAR technology, closed doors, attendance, and so on. Finally, u_{itH} is the error term. In all regressions, to take into account unobservable correlated factors within a match, we cluster standard errors at the match level.

We are interested in the coefficients β_1 and β_2, which measure how the probability of awarding a penalty (or a red card or yellow card) changes based on the referee's previous decisions. Controlling for variables that capture the likelihood of a decision – such as the current score, the quality of the teams and the tactics they are adopting – if referees assess each match episode independently, case by case, β_1 and β_2 should both be zero. In contrast, if $\beta_1 < 0$ and $\beta_2 > 0$, this would provide evidence of a compensatory preference by the referee, whereby previous decisions in favor of the home team are likely to be followed by decisions in favor of the away team, and vice versa.

We also estimate an analogous equation for the away team, Y_{itA}:

$$Y_{itA} = \gamma_0 + \gamma_1 \left(\sum_{\tau=1}^{t-1} Y_{i\tau A} \right) + \gamma_2 \left(\sum_{\tau=1}^{t-1} Y_{i\tau H} \right) + \gamma_3 ScoreDiff_{it-1} + \gamma_4 Cum.\boldsymbol{X_{it-1}} + \gamma_5 \boldsymbol{W_i} + \gamma_6 \boldsymbol{Z_i} + v_{itA}$$

in which we are interested in the coefficients γ_1 and γ_2, which can be interpreted analogously to β_1 and β_2.

5.1. Referee Decisions on Penalties

Before conducting our regression analyses, we provide additional preliminary evidence by showing, in Figure 1, the probability of awarding a penalty to the home team (solid line) and the away team (dashed line) as a function of the difference in the home and away penalties previously awarded. Clearly, the probability of awarding a (further) penalty to the home team is notably high when one or two penalties have been awarded to the away team, while it decreases when one or two penalties have already been awarded to the home team. Conversely, the probability of awarding a penalty to the away team is relatively low when one or two penalties have been awarded to the away team, but it increases when the home team has previously been awarded one or two penalties.

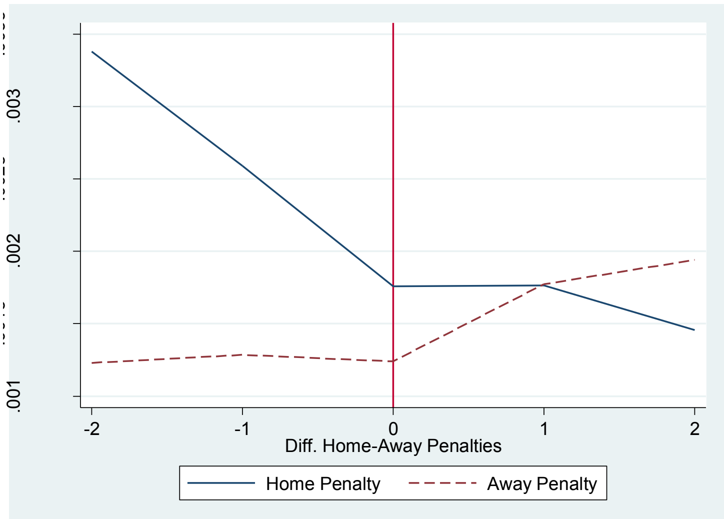

Figure 1. The Probability of Awarding a Penalty to the Home and Away Teams

Although the graphical analysis provides preliminary insights, we proceed with a regression analysis that allows us to take into account multiple factors that simultaneously affect the probability of awarding a penalty. In Table 3, we report the estimates of a Linear Probability Model for the probability of awarding a penalty to the home team at a given minute *t* of the match. To enhance the interpretability of the coefficients, we multiply them by 100: in this way, each coefficient reflects the change in the probability of awarding a penalty over the entire match, rather than within a one-minute frame. In all regressions, standard errors are adjusted for heteroskedasticity and clustered at the match level.

In column (1) we show that the probability of awarding a penalty to the home team is negatively related to the number of penalties awarded to the home team up to minute t-1: specifically, the probability decreases by 2.4 percentage points (t-stat=–2.38). On the other hand, the probability of awarding a penalty increases by 4.7 p.p. if the away team has been previously awarded a penalty (t-stat=3.32). In this specification, we also control for the *Score Difference* at t–1. If the home team is ahead by one goal, our estimates show that the probability of obtaining a penalty is 3.1 p.p. lower, likely because the home team is attacking less frequently. Some studies (for example, Anderson and Pierce, 2009; Considine *et al.*, 2024) have interpreted the negative coefficient on the current score as evidence of a compensatory tendency by referees towards teams that are losing. While this may be the case, it is difficult to separate this effect from the fact that losing teams tend to attack more frequently and are thus more often present in the penalty area.

We also control for the number of *Shots* of home and away teams up to time t–1, as an additional proxy for offensive and defensive strategies adopted by the two teams. We find that more shots by a team increase its probability of being awarded a penalty, and vice versa. Furthermore, we control for the quality of the opposing teams, by using the home team's probability of winning, based on betting odds. As expected, we find that the stronger the home team, the higher its probability of receiving a penalty.

In column (2) of Table 3, we additionally control for the cumulated number of fouls: as expected, more fouls committed by the away team increase the probability of a penalty for the home team and vice versa. The coefficients on cumulated penalties of home and away teams remain similar: –4.4 p.p. (t-stat=–4.15) for an additional penalty previously awarded to the home team and +4.3 p.p. (t-stat=3.04) for an additional penalty awarded to the away team.

In column (3), we use directly the difference in cumulated penalties between the home team and the away team, labeled "Cumulated Penalties Home-Away", rather than including the two variables separately. We find that a positive difference (one penalty more for the home team) decreases the probability of awarding a penalty to the home team by 4.4 p.p. (t-stat=–4.8), while a negative difference (one penalty more for the away team) increases it.

In column (4) of Table 3, instead of using the cumulated shots and fouls up to time t–1, we use the cumulative shots and fouls in the last 15 minutes, to consider a different proxy for the tactics adopted by the two teams in the latest stage of the match. Our coefficients of interest remain very similar. Finally, in column (5), we estimate the model of column (2) but we include referee and season fixed effects: we find that home teams are less likely to be awarded a penalty if they have already received one (–6.4 p.p.), while they are more likely to receive a penalty if the away team has been awarded one previously (+3.8 p.p.).

Table 3. The Probability of Awarding a Penalty to the Home Team. Linear Probability Model

	(1) Penalty Home	(2) Penalty Home	(3) Penalty Home	(4) Penalty Home	(5) Penalty Home
Cum. Penalties Home	-0.024**	-0.044***		-0.027***	-0.064***
	(0.010)	(0.011)		(0.010)	(0.011)
Cum. Penalties Away	0.047***	0.043***		0.053***	0.038***
	(0.014)	(0.014)		(0.014)	(0.014)
Cum. Penalties Home-Away			-0.044***		
			(0.009)		
Score Diff. Home-Away	-0.031***	-0.030***	-0.030***	-0.024***	-0.029***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Prob. Home Team Win	0.228***	0.210***	0.210***	0.241***	0.211***
	(0.017)	(0.018)	(0.018)	(0.017)	(0.018)
Cum. Shots Home	0.008***	0.002***	0.002***		0.003***
	(0.001)	(0.001)	(0.001)		(0.001)
Cum. Shots Away	-0.002**	-0.007***	-0.007***		-0.006***
	(0.001)	(0.001)	(0.001)		(0.001)
Cum. Fouls Home		-0.010***	-0.010***		-0.010***
		(0.001)	(0.001)		(0.001)
Cum. Fouls Away		0.024***	0.024***		0.024***
		(0.001)	(0.001)		(0.001)
Cum. Shots Home (Last 15m)				-0.002	
				(0.002)	
Cum. Shots Away (Last 15m)				-0.010***	
				(0.002)	
Cum. Fouls Home (Last 15m)				-0.019***	
				(0.002)	
Cum. Fouls Away (Last 15m)				0.098***	
				(0.003)	
Constant	0.036***	0.020**	0.020**	-0.043***	-0.007
	(0.009)	(0.009)	(0.009)	(0.010)	(0.015)
Referee and Season F.E.	NO	NO	NO	NO	YES
Observations	2119986	2119986	2119986	2119986	2119986
Adjusted R^2	0.000	0.000	0.000	0.001	0.001

Notes: The table reports OLS estimates. The dependent variable is *Penalty Home*. Standard errors (reported in parentheses) are adjusted for heteroskedasticity and clustered at the match level. The symbols ***, ** and * indicate that coefficients are statistically significant at the 1%, 5%, and 10% levels, respectively.

To assess the robustness of our results, we have run several alternative specifications: (a) using the cumulated penalties awarded to home and away teams in the last 15 minutes (the latter variables turn out to have a stronger impact); (b) including controls for the cumulated corners and offsides of home and away teams; (c) including league fixed effects; (d) controlling for 10 dummy variables, each representing a 10-minute interval of the match (1-10 minutes, 11-20, etc.);[7] (e) controlling for the length of the injury time awarded in both halves; (f) using the difference in Elo ratings instead of the probability of winning for the home team; (g) controlling for cumulated red and yellow cards awarded to both teams; (h) using a Probit estimator instead of a Linear Probability Model. In all these specifications – which are not reported here to avoid cluttering the paper – we obtain very similar results.

[7] The probability of awarding a penalty to the home team decreases as the game progresses, especially when the score is tied.

Table 4 shows the estimates for the probability of awarding a penalty to the away team. We estimate the same specifications as those in Table 3, but we report only the coefficients on cumulative penalties awarded up to time t-1. We find the same pattern of behavior by referees in balancing their decisions: the probability to award a penalty to the away team increases by approximately 2-3 percentage points if a penalty has previously been awarded to the home team, whereas it decreases of about 2-3 p.p. if the away team has previously benefited from a penalty.

Table 4. The Probability of Awarding a Penalty to the Away Team. Linear Probability Model

	(1) Penalty Away	(2) Penalty Away	(3) Penalty Away	(4) Penalty Away	(5) Penalty Away
Cum. Penalties Home	0.032***	0.027***		0.032***	0.022**
	(0.010)	(0.010)		(0.010)	(0.010)
Cum. Penalties Away	-0.012	-0.027***		-0.014	-0.049***
	(0.010)	(0.011)		(0.010)	(0.011)
Cum. Pen. Home-Away			0.027***		
			(0.008)		
Observations	2119986	2119986	2119986	2119986	2119986
Adjusted R^2	0.000	0.000	0.000	0.001	0.000

Notes: The Table reports OLS estimates. The dependent variable is *Penalty Away*. We estimate the same specifications as those in Table 3. Standard errors (reported in parentheses) are adjusted for heteroskedasticity and clustered at the match level. The symbols ***, ** and * indicate that coefficients are statistically significant at the 1%, 5%, and 10% levels, respectively.

5.2. Decisions on Red Cards

Similarly to the approach used for penalty kicks, Figure 2 illustrates the probabilities of issuing a red card to a home team player (solid line) or an away team player (dashed line) as a function of the red card differential between the two teams at a given time. It clearly emerges that if the away team has previously received two more red cards than the home team (a difference of –2), this results in a higher probability of the home team receiving a red card. When there is no difference in the number of red cards awarded to the two teams, the probability of receiving a red card is similar for both teams (although the away team is slightly more likely to receive one), whereas the probability of the away team receiving a red card increases significantly if the home team has previously received one or two more red cards.

Table 5 presents the OLS estimates for the probability of a referee issuing a red card to the home team at a given minute t. As in previous tables, the coefficients are scaled by 100 for improve readability. Since red cards are strongly correlated to prior yellow cards, all the specifications include controls for yellow cards awarded previously.

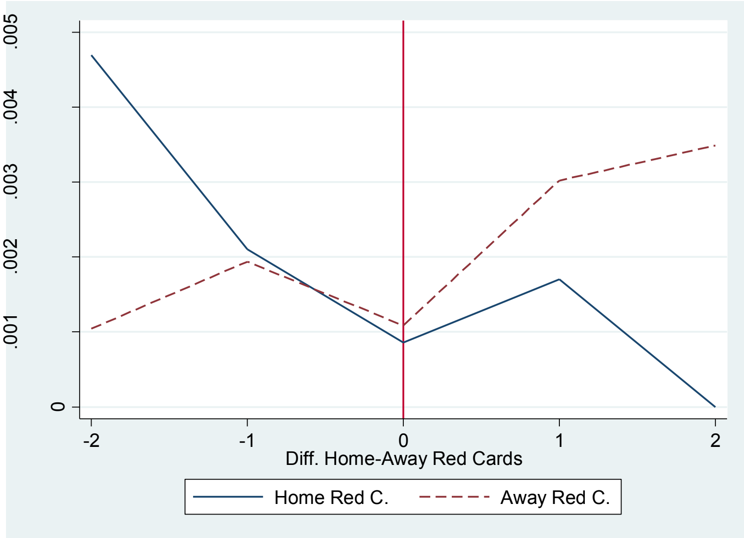

Figure 2. The Probability of a Red Card for a Player of the Home and the Away Team

The probability of a home team player receiving a red card decreases by 5.3 p.p. if the home team has already been issued a red card (*t*-stat=–3.35). Conversely, if the away team has previously been issued a red card, the likelihood of a home team player receiving a red card increases significantly by 6.3 p.p. (t-stat=3.64).

When we include controls for the cumulative fouls committed by both teams up to t–1, the results remain consistent, suggesting that the referees' tendency to balance their decisions persists even when accounting for the teams' overall level of aggressiveness (col. 2, Table 5).

Table 5. The Probability of Showing a Red Card to a Home Team Player. Linear Probability Model

	(1) Red Card Home	(2) Red Card Home	(3) Red Card Home	(4) Red Card Home	(5) Red Card Home
Cum. Red Cards Home	-0.053***	-0.052***		-0.045***	-0.072***
	(0.016)	(0.016)		(0.016)	(0.016)
Cum. Red Cards Away	0.063***	0.062***		0.060***	0.056***
	(0.017)	(0.017)		(0.017)	(0.017)
Cum. Red Cards Home-Away			-0.058***		
			(0.013)		
Observations	2119986	2119986	2119986	2119986	2119986
Adjusted R^2	0.001	0.001	0.001	0.001	0.001

Notes: The Table reports OLS estimates. The dependent variable is *Home Red Card*. We estimate the same specifications as those in Table 3 but include the cumulative yellow cards among controls. Standard errors (reported in parentheses) are adjusted for heteroskedasticity and clustered at the match level. The symbols ***, ** and * indicate that coefficients are statistically significant at the 1%, 5%, and 10% levels, respectively.

Overall, the estimates of referees' tendency to balance previous decisions in favor of or against the two teams appear robust: across all models, cumulative red cards for the home team consistently exhibit a negative effect, while cumulative red cards for the away team show a positive effect on the likelihood of the home team receiving a red card.

In Table 6 we estimate the probability of issuing a red card to an away team player. We use the same specifications as those used in Table 5, reporting only the coefficients on the variables "Cumulative Red

Cards". Consistent with our hypothesis, we find that, across all model specifications, the probability of an away team player receiving a red card increases by 8 p.p. if the home team has previously been issued a red card, while it decreases by 6.9 p.p. if the away team has already received a red card. These results remain robust when we incorporate controls for fouls and shots (columns 2 and 4, Table 6), as well as when we include referee and season fixed effects (column 5, Table 6). This consistency provides further support to the hypothesis that referees balance their decisions based on prior events, demonstrating a clear pattern of compensatory behavior in refereeing.

Table 6. The Probability of Showing a Red Card to an Away Team Player. Linear Probability Model

	(1) Red Card Away	(2) Red Card Away	(3) Red Card Away	(4) Red Card Away	(5) Red Card Away
Cum. Red Cards Home	0.080***	0.078***		0.080***	0.070***
	(0.023)	(0.023)		(0.023)	(0.023)
Cum. Red Cards Away	-0.069***	-0.067***		-0.063***	-0.088***
	(0.016)	(0.016)		(0.016)	(0.016)
Cum. Red Cards Home-Away			0.072***		
			(0.014)		
Observations	2119986	2119986	2119986	2119986	2119986
Adjusted R^2	0.001	0.001	0.001	0.001	0.001

Notes: The Table reports OLS estimates. The dependent variable is *Red Card Away*. We estimate the same specifications as those in Table 3 but include the cumulative yellow cards among controls. Standard errors (reported in parentheses) are adjusted for heteroskedasticity and clustered at the match level. The symbols ***, ** and * indicate that coefficients are statistically significant at the 1%, 5%, and 10% levels, respectively.

5.3. Decisions on Yellow Cards

A similar pattern emerges for another critical category of refereeing decisions: yellow cards. Figure 3 illustrates the probability of each team receiving a yellow card as a function of the yellow card differential between the home and away teams at a given time. In this case, even more so than for penalties and red cards, the differential significantly influences the likelihood of receiving a yellow card.

This effect is confirmed in Tables 7 and 8, which report the estimates of the probability of receiving a yellow card at minute t for the home team and the away team, respectively: referees tend to show more yellow cards to the team that previously received fewer yellow cards, and vice versa.

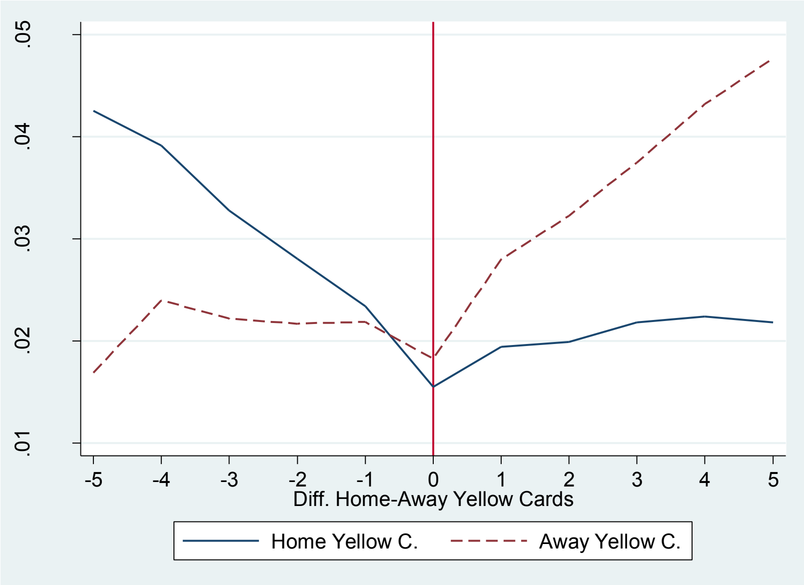

Figure 3. The Probability of Receiving a Yellow Card for the Home and the Away Team

Table 7. The Probability of Issuing a Yellow Card to a Home Team Player. Linear Probability Model

	(1) Yellow Card Home	(2) Yellow Card Home	(3) Yellow Card Home	(4) Yellow Card Home	(5) Yellow Card Home
Cum. Yellow Cards Home	-0.044***	-0.277***		-0.008	-0.359***
	(0.013)	(0.015)		(0.012)	(0.015)
Cum. Yellow Cards Away	0.370***	0.256***		0.472***	0.196***
	(0.012)	(0.014)		(0.012)	(0.014)
Cum. Yellow Cards Home-Away			-0.266***		
			(0.011)		
Observations	2119986	2119986	2119986	2119986	2119986
Adjusted R^2	0.003	0.004	0.004	0.003	0.005

Notes: The Table reports OLS estimates. The dependent variable is *Yellow Card Home*. We estimate the same specifications as those in Table 3. Standard errors (reported in parentheses) are adjusted for heteroskedasticity and clustered at the match level. The symbols ***, ** and * indicate that coefficients are statistically significant at the 1%, 5%, and 10% levels, respectively.

Table 8. The Probability of Issuing a Yellow Card to an Away Team Player. Linear Probability Model

	(1) Yellow Card Away	(2) Yellow Card Away	(3) Yellow Card Away	(4) Yellow Card Away	(5) Yellow Card Away
Cum. Yellow Cards Home	0.414***	0.282***		0.524***	0.221***
	(0.014)	(0.015)		(0.013)	(0.015)
Cum. Yellow Cards Away	-0.070***	-0.339***		-0.027**	-0.424***
	(0.013)	(0.015)		(0.011)	(0.015)
Cum. Yellow Cards Home-Away			0.312***		
			(0.011)		
Observations	2119986	2119986	2119986	2119986	2119986
Adjusted R^2	0.003	0.004	0.004	0.003	0.005

Notes: The Table reports OLS estimates. The dependent variable is *Yellow Card Away*. We estimate the same specifications as those in Table 3. Standard errors (reported in parentheses) are adjusted for heteroskedasticity and clustered at the match level. The symbols ***, ** and * indicate that coefficients are statistically significant at the 1%, 5%, and 10% levels, respectively.

5.4. Robustness Checks: Estimates Using 10 Frames per Match

In this Section, we divide each match in our dataset into 10 frames of approximately 10 minutes each, rather than using 100 frames per match. This serves as an additional robustness check to assess whether the chosen frame structure introduces any potential bias.

For each frame, we calculate the total number of decisions awarded to each team (penalties, red cards, yellow cards), along with other events such as shots, and fouls. Regarding the *Score Difference* between the home team and the away team, we consider the difference observed at the beginning of each frame.

We then estimate the probability of awarding a penalty, a red card, and a yellow card to the home and away team in a given frame, using specification (5) from Table 3, which includes referee and season fixed effects. The results, presented in Table 9, closely align with previous findings. Controlling for the score difference between the two teams, we find that the probability of awarding a penalty to one team is negatively related to the number of penalties already awarded to the same team and positively related to those awarded to the opposing team (columns 1 and 2). Similar patterns emerge for referee decisions regarding red cards (columns 3-4) and yellow cards (columns 5-6).

Table 9. Estimates with Matches Split in 10 Frames

	(1) Penalty Home	(2) Penalty Away	(3) Red Card Home	(4) Red Card Away	(5) Yellow Card Home	(6) Yellow Card Away
Cum. Penalties Home	-0.057***	0.030***				
	(0.011)	(0.011)				
Cum. Penalties Away	0.040***	-0.037***				
	(0.015)	(0.011)				
Cum. Red Cards Home			-0.041**	0.086***		
			(0.017)	(0.024)		
Cum. Red Cards Away			0.050***	-0.039**		
			(0.018)	(0.017)		
Cum. Yellow Cards Home					-0.313***	0.170***
					(0.015)	(0.016)
Cum. Yellow Cards Away					0.159***	-0.360***
					(0.015)	(0.015)
Score Diff. Home-Away	-0.029***	0.022***	-0.019***	0.009***	-0.101***	-0.072***
	(0.003)	(0.003)	(0.002)	(0.003)	(0.010)	(0.011)
Cum. Shots Home	0.006***	-0.003***	0.001	0.001	0.019***	0.036***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)
Cum. Shots Away	-0.002**	0.005***	0.003***	0.001	0.033***	0.014***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)	(0.004)
Cum. Fouls Home	-0.001	0.004***	0.012***	0.005***	0.134***	0.063***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)	(0.005)
Cum. Fouls Away	0.007***	0.002**	0.004***	0.013***	0.045***	0.144***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)	(0.005)
Prob. Home Team Win	0.232***	-0.150***	-0.015	0.008	-0.989***	0.460***
	(0.017)	(0.014)	(0.011)	(0.013)	(0.053)	(0.057)
Referee and Season FE	YES	YES	YES	YES	YES	YES
Observations	214140	214140	214140	214140	214140	214140
R^2	0.005	0.004	0.009	0.009	0.046	0.043

Notes: The Table reports OLS estimates. The dependent variables are indicated at the top of each column. We estimate the same specification as in column (5) of Table 3. Standard errors (reported in parentheses) are adjusted for heteroskedasticity and clustered at the match level. The symbols ***, ** and * indicate that coefficients are statistically significant at the 1%, 5%, and 10% levels, respectively.

6. Exploring Heterogeneous Effects

In this Section, we investigate whether the compensatory tendencies of referees vary across different circumstances and contexts, to shed light on potential underlying mechanisms.

First, we analyze whether referees' behavior changes when a match is played with a crowd or behind closed doors. Specifically, we analyze whether crowd pressure influence referees' compensatory tendency. Since in the previous sections we have shown that the referees tend to compensate both the home team and the away team after an own discretionary decision in favor of the opposing team, we expect that the crowd presence does not significantly influences referees' behavior.

In Panel (a) of Table 10 (columns 1 and 2) we examine the probability of awarding a penalty to the home team, replicating specification (5) of Table 3, but using the *Difference in Cumulated Penalties Home-Away* as a synthetic indicator of the compensatory tendency of referees. In column (1) we estimate on the sample of matches with a crowd, while in column (2) we use the matches played behind closed doors during the Covid-19 pandemic. We find that the compensatory behavior of referees is strong and significant with a crowd (–5.5 p.p.), but becomes even more pronounced in the absence of spectators (–8.3 p.p.). A similar pattern emerges for red card decisions in columns (3) and (4), and for yellow card decisions (columns 5 and 6). These findings suggest that the uncovered behavior of referees reflects intrinsic preferences, but that the presence of the crowd can affect its magnitude.

In Panel (b) of Table 10, we analyze whether referees' behavior differs between matches in which VAR technology has been used to assist refereeing decisions with respect to situations in which VAR was not available.[8] The compensatory tendencies turn out to be present in both contexts, but we find that for penalties the magnitude of the effects are attenuated with the availability of VAR technology (–5.6 vs. –6.1 p.p.).

The difference is more pronounced for red cards: referees' compensatory tendencies appear to be considerably attenuated with the introduction of VAR, decreasing from –7.1 without VAR to –4.4 with VAR. For yellow cards – which are not subjected to VAR review – as expected, we do not observe any significant change (–0.29 vs –0.25). Overall, while compensatory tendencies persist with VAR, their magnitude is somewhat reduced by the technology.

We next examine whether the referees' age or experience – two highly correlated variables in our sample – affects their compensatory behavior. On the one hand, older referees with greater experience, may have developed the ability to judge each episode on its merit, making them more resistant to inequity aversion. On the other hand, older referees, being less concerned with career advancements, might exhibit stronger inequity aversion, even at the cost of occasional mistakes. Our estimates in Panel (c) – in which we split the sample distinguishing referees with age below the median (39 years) from those above it – show that, for penalties and red cards, younger referees tend to be slightly more inclined to compensate teams for their previous decisions with respect to their older counterparts (–6.6 vs –4.4). The same pattern emerges for red

[8] Notice that VAR can only assist referees in specific situations: goals decisions, penalties, red cards and identifying the correct player for disciplinary sanctions.

cards (–7.2 vs. –3.9, respectively). Similar results emerge when using referee experience, measured as the number of years since their debut up to match i, instead of age.

Table 10. Heterogeneous Effects in Compensatory Tendency

Panel (a). Crowd vs. Closed Doors

	Penalty Home		Red Card Home		Yellow Card Home	
	(1) Crowd	(2) Closed	(3) Crowd	(4) Closed	(5) Crowd	(6) Closed
Cum. Pen. Home-Away	-0.055***	-0.083***				
	(0.010)	(0.027)				
Cum. Red Home-Away			-0.056***	-0.090**		
			(0.013)	(0.043)		
Cum. Yellow Home-Away					-0.269***	-0.341***
					(0.011)	(0.034)
Observations	1896444	223542	1896444	223542	1896444	223542

Panel (b). VAR vs. No VAR

	Penalty Home		Red Card Home		Yellow Card Home	
	(1) No VAR	(2) VAR	(3) No VAR	(4) VAR	(5) No VAR	(6) VAR
Cum. Pen. Home-Away	-0.061***	-0.055***				
	(0.013)	(0.012)				
Cum. Red Home-Away			-0.071***	-0.044**		
			(0.017)	(0.018)		
Cum. Yellow Home-Away					-0.298***	-0.254***
					(0.015)	(0.015)
Observations	1035639	1084347	1035639	1084347	1035639	1084347

Panel (c). Referee's Age

	Penalty Home		Red Card Home		Yellow Card Home	
	(1) Age<=39	(2) Age>39	(3) Age<=39	(4) Age>39	(5) Age<=39	(6) Age>39
Cum. Pen. Home-Away	-0.066***	-0.044***				
	(0.014)	(0.013)				
Cum. Red Home-Away			-0.072***	-0.039**		
			(0.019)	(0.017)		
Cum. Yellow Home-Away					-0.273***	-0.279***
					(0.016)	(0.016)
Observations	917334	917334	917334	917334	917334	917334

Notes: The Table reports OLS estimates. The dependent variables are reported at the top of each column. We estimate specification (5) of Table 3 including league and season fixed effects. Standard errors (reported in parentheses) are adjusted for heteroskedasticity and clustered at the match level. The symbols ***, ** and * indicate that coefficients are statistically significant at the 1%, 5%, and 10% levels, respectively.

Furthermore, we explored whether compensatory tendencies vary between the first and second halves of the match. We also examined differences between more contested matches – where the home team's probability of winning falls in the second and third quartiles – and less contested matches. However, no significant differences were observed in either case.

7. The Decision on Injury Time: Giving a Chance to Any Team Behind

As discussed in Section 2, several studies in the literature on social pressure in stadiums – such as those by Garicano, Palacios-Huerta, and Prendergast (2005) and Dohmen (2008), among others – have shown that referees are influenced by crowd pressure. Specifically, they are inclined to allocate additional injury time at the end of a match when the home team is behind, increasing its chances of equalizing.

However, if crowd pressure were the sole driver of referees' behavior, one would also expect them to reduce injury time when the home team is leading. In contrast, simple descriptive statistics (see Figure 4) reveal that referees tend to allow more injury time not only when the home team is trailing by one goal (+0.44) but also when the away team is behind (+0.22), compared to when the score is tied. While social pressure is undoubtedly a key factor in determining injury time, this evidence suggests the influence of additional forces.

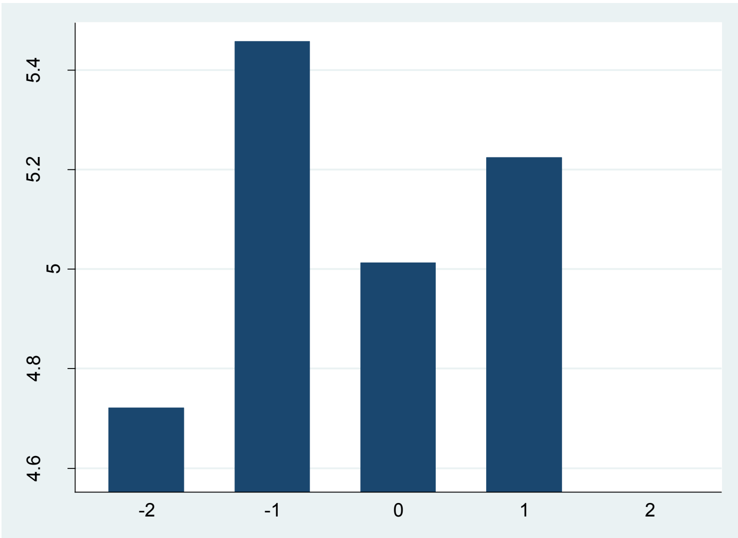

Figure 4. Injury Time as a Function of the Score Difference at the 90th Minute

If referees aim to maintain neutrality and treat teams fairly, they may be inclined to allocate more injury time when either team is behind. In this Section we conduct an econometric analysis of the additional time allowed by referees at the end of regular time, accounting for several relevant factors.

Our analysis uses match-level data, focusing on matches with a score difference of –1, 0, and 1 at the end of regular time (90th minute), since in these situations the score can plausibly be changed in the few additional minutes allowed. The dataset includes a total of 13,810 matches.

Table 11 presents descriptive statistics for this sample. All the variables (substitutions, red and yellow cards, injuries, fouls, etc.) are calculated by summing up all such events for both teams in the second half, since soccer rules dictate that referees must add time to account for delays caused by these occurrences.

On average, injury time is 5.2 minutes. In 27% of the matches, the home team is behind; in 33% the away team is behind; and in 40% the score is tied. In the second half, there are on average 5.9 substitutions,[9] 0.13 red cards, 2.4 yellow cards, and 0.14 penalties.

Table 11. Descriptive Statistics for the Injury Time Analysis

Variable	Mean	Std. Dev.	Min	Max	Obs
Injury Time	5.2017	1.6652	0	18	13810
Assigned Injury Time	4.2641	1.3727	1	15	10673
Home Behind	0.2648	0.4413	0	1	13813
Draw	0.3994	0.4898	0	1	13813
Away Behind	0.3358	0.4723	0	1	13813
# Substitutions	5.9375	1.7146	1	10	13813
# Injuries	0.4250	0.6435	0	4	13813
#VAR Reviews	0.1024	0.3456	0	6	13813
# Red Cards	0.1265	0.3670	0	3	13813
# Yellow Cards	2.4352	1.5950	0	11	13813
# Penalties	0.1387	0.3738	0	3	13813
# Fouls	11.4140	3.6025	1	29	13813
# Shots	11.6526	3.6412	1	32	13813
# Goals	1.1444	1.0767	-1	7	13813
Prob. Home Team Win	0.4309	0.1668	0.0421	0.9152	13808

Notes: The variables are measured at the end of regular time (90th minute). The number of events (e.g., red cards, yellow cards) represents the total for both the home and away teams and is based solely on occurrences during the second half.

We estimate the following equation:

$$Injury\ Time_i = \delta_0 + \delta_1 HomeBehind_i + \delta_2 AwayBehind_i + \delta_3 \boldsymbol{X}_i + \delta_4 W_i + u_i$$

where $Injury\ Time_i$ represents the additional time awarded by the referee in match i, *Home Behind* is a dummy equal to one when the home team is behind by one goal (and zero otherwise), while *Away Behind* is equal to one when the away team is behind by one goal (and zero otherwise). The vector $\boldsymbol{X}$ includes a set of variables that could influence *Injury Time* such as the number of substitutions, red cards, yellow cards, injuries, VAR reviews, penalties, goals, and so on. Additionally, we control for a variable W that measures the quality of the competing teams, based on either betting odds or Elo ratings. In all specifications, we adjust Standard Errors for heteroskedasticity and for clustering at the league and season level.

Our OLS estimates are reported in Table 12. In the first column, we include only the two dummies, *Home Behind* and *Away Behind*. We find that – compared to when the score is tied – referees add 0.44 minutes (t-stat=12.2) when the home team is behind, but also add on average 0.21 minutes when the away team is behind (t-stat=6.3). This finding contrasts sharply with the explanation based on social pressure, which suggests that the crowd would pressure referees to shorten injury time when the home team is ahead by one goal. Importantly, the pressure to shorten injury time when the home team is leading should be stronger than the pressure to lengthen it when the home team is trailing. In the former case two points are at stake (as a win

[9] Before the Covid-19 pandemic (up to the mid 2019-20 season), each team was allowed 3 substitutions per match; after the pandemic, this was increased to 5 substitutions per team.

could turn into a draw), while in the latter case only one point is realistically at stake (providing the opportunity to turn a loss into a draw).

In contrast, our finding suggests that referees may seek to be perceived as fair by players and external observers, providing the trailing team—whether home or away—an opportunity to catch up.

In column (2) we control for the number of substitutions, injuries and VAR reviews: all these variables, as expected, increase injury time, with VAR reviews having the largest effect (1.19 minutes per episode). However, the magnitude and statistical significance of the two dummies, *Home Behind* and *Away Behind*, are only slightly diminished.

In the subsequent columns of Table 12, we additionally control for red and yellow cards, penalties and fouls (column 3), for shots, goals scored and the probability of the home team winning (column 4) and for referee and season fixed effects (column 5). In all specifications, we find that when either the home team or the away team is behind, referees allow additional injury time, more so when the home team is behind (+0.38 in column 5, *t*-stat=9.9), suggesting a role for social pressure, but also when the away team is behind (+0.16, *t*-stat=4.40).

Together, these results suggest that while crowd pressure is an important factor, referees' inequity aversion and pursuit of impartiality also play a significant role in their decisions.

In the analysis of Table 12, we used the effective injury time, as this variable is available for the entire sample. Ideally, one should use the injury time assigned by the referee at the 90th minute. However, this variable is available for a smaller sample, collected from the website www.diretta.it, as the assigned injury time was not reported prior to the 2013-14 season. For contested matches (score differences of –1, 0, 1 at 90th minute) we have 10,673 observations, compared to 13,810 in the full sample. The correlation between the two measures of injury time is 0.89. In the appendix, Table A2 reports estimates using the *Assigned Injury Time* as the dependent variable: these results are nearly identical to those of Table 12.

Our results remain consistent if we also consider matches with score differences of –2 and +2. When the absolute score difference is 2, referees tend to shorten injury time, likely because the chances of altering the final outcome of the match are considered minimal. The results are also robust when we use a Poisson estimator, instead of an OLS, accounting for the fact that injury time takes non-negative integer values.

We further investigate whether the effects differ when matches are played with a crowd or behind closed doors. For matches with a crowd, the time added when the home team is behind (+0.37) is more than twice that when the away team is behind (+0.15), but both are statistically significant. Behind closed doors, the difference between the injury time added for the home team (+0.49 minutes) and for the away team (+0.35 minutes) narrows (both coefficients are statistically significant).

Additionally, we find almost no difference in the magnitude of the coefficients on *Home Behind* and *Away Behind* when VAR technology is available.

Table 12. Injury Time as a Function of Score Difference. OLS estimates

	(1) Injury Time	(2) Injury Time	(3) Injury Time	(4) Injury Time	(5) Injury Time
Home Behind	0.445***	0.377***	0.360***	0.339***	0.382***
	(0.036)	(0.037)	(0.037)	(0.041)	(0.038)
Away Behind	0.212***	0.158***	0.151***	0.112**	0.160***
	(0.034)	(0.033)	(0.032)	(0.035)	(0.036)
# Substitutions		0.232***	0.227***	0.221***	0.076***
		(0.039)	(0.038)	(0.037)	(0.016)
# Injuries		0.422***	0.419***	0.417***	0.396***
		(0.037)	(0.036)	(0.036)	(0.030)
#VAR Reviews		1.188***	1.095***	1.079***	0.790***
		(0.103)	(0.097)	(0.095)	(0.076)
# Red Cards			0.217***	0.217***	0.307***
			(0.043)	(0.043)	(0.037)
# Yellow Cards			0.090***	0.088***	0.081***
			(0.019)	(0.019)	(0.012)
# Penalties			0.143**	0.071	0.115*
			(0.050)	(0.047)	(0.044)
# Fouls			-0.059***	-0.059***	-0.012**
			(0.009)	(0.009)	(0.004)
# Shots				-0.023***	-0.021***
				(0.006)	(0.004)
# Goals				0.121**	0.110***
				(0.039)	(0.027)
Prob. Home Team Win				0.153*	0.230**
				(0.067)	(0.067)
Constant	5.013***	3.370***	3.821***	3.961***	3.588***
	(0.119)	(0.222)	(0.233)	(0.221)	(0.137)
Referee and Season FE	NO	NO	NO	NO	YES
Observations	13810	13810	13810	13805	13805
Adjusted R^2	0.011	0.175	0.194	0.201	0.369

Notes: The Table reports OLS estimates. The dependent variable is *Injury time*. The variables *Home Behind* and *Away Behind* are evaluated at the end of regular time (90th minute). The number of events (substitutions, red cards, yellow cards, and so on) refer to the second half. Standard errors (reported in parentheses) are adjusted for heteroskedasticity and clustered at the league and season level. The symbols ***, ** and * indicate that coefficients are statistically significant at the 1%, 5%, and 10% levels, respectively.

8. Concluding Remarks

In this paper, using minute-by-minute observations from approximately 21,400 soccer matches, we have investigated how referees' disciplinary decisions during a match are influenced by their prior decisions in favor of or against each team.

Although referees are expected to make decisions independently on each episode, we have developed a theoretical model suggesting that if referees are motivated by a preference to treat both teams fairly and experience discomfort when creating an imbalance between the teams – or if they desire to be perceived as fair by players, fans, experts and others – they might raise the evidence threshold to award a second penalty (or a red or yellow card) to the same team. Conversely, they could lower the threshold to award a penalty to the opposing team.

Our econometric analysis has explored whether the probability of a referee awarding a penalty or issuing a red or a yellow card at a given minute of the match is influenced by similar decisions against the same team or the opposing team in earlier phases of the match. Throughout our analysis, we control for a large set of variables that reflect the defensive and attacking tactics that teams are adopting at a given time and for the quality of the teams involved.

We have shown that the referee's decision to award a penalty is negatively correlated with penalties awarded to the same team (about –6 p.p.), while it is positively correlated with penalties awarded to the opposing team (+3.8 p.p.). A similar pattern is observed for red and yellow card decisions.

Our results remain robust when we include a wide range of controls and conduct several sensitivity checks. When examining whether these effects vary under different conditions (e.g., matches played behind closed doors, the use of VAR technology, and referee experience), we find that the compensatory tendencies of referees do not change significantly. This suggests that their behavior is likely driven by intrinsic preferences, rather than being shaped by external factors.

In the final part of the paper we analyzed referees' decisions regarding injury time. On the one hand, we confirm that the crowd pressure does affect referees, as they tend to allow more injury time when the home team is behind. However, in stark contrast to explanations based purely on crowd pressure, we find that, instead of shortening the injury time when the home team is leading, referees tend to extend it. This evidence further supports the notion that referees aim to treat teams fairly and are averse to being decisive for the final outcome.

The implications of our findings extend beyond soccer. In several kinds of subjective assessment or evaluations in real-world settings – such as employee evaluations carried out by managers, student assessments by teachers, judicial decisions in Courts, and so on – decision-makers may not only be influenced by extrinsic incentives or social pressure, but may also be motivated by a desire to treat all parties fairly and to be perceived as fair. Thus, in these settings, decisions could be shaped by a compensatory tendency, rather than solely by rational, unbiased, and unemotional evaluation.

References

Anderson, K. J., and Pierce, D. A. (2009). "Officiating bias: The effect of foul differential on foul calls in NCAA basketball", *Journal of Sports Sciences*, 27(7), 687-694.

Bénabou, R., and Tirole, J. (2006). "Incentives and prosocial behavior", *American Economic Review*, 96(5), 1652-1678.

Bolton, G., and Ockenfels, A. (2000), "ERC: A Theory of Equity, Reciprocity, and Competition", *American Economic Review*, 166-193.

Charness, G., and Rabin, M. (2002), "Understanding Social Preferences with Simple Tests", *Quarterly Journal of Economics*, 117, 3, 817-869.

Chen, D. Moskowitz, T., and Shue, K., (2016), "Decision Making Under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires", *Quarterly Journal of Economics*, 131, 3, 1181–1242.

Considine, J., Eakins, J., Horgan, P., and Weir, C. (2024), "Compensating Tendencies in Disciplinary Sanctions: The Case of Hurling", *Journal of Sports Economics*, 15270025241245640.

Dawson, P. and Dobson, S. (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 Wilson, J. (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 (Statistics in Society)*, 170, pp. 231-250.

Dohmen, T. (2008). “The influence of social forces: Evidence from the behavior of football referees”. *Economic Inquiry*, 46(3), 411-424.
Dohmen, T., and Sauermann, J. (2016), “Referee bias”, *Journal of Economic Surveys*, 30(4), 679-695.
Fehr, E., and Schmidt, K. (1999), “A Theory of Fairness, Competition, and Cooperation”, *Quarterly Journal of Economics*, 114, 817-868.
Garicano, L., Palacios-Huerta, I. and Prendergast, C. (2005). “Favoritism under Social Pressure”. *Review of Economics and Statistics*, 87, 208-16.
Haynes, G., and Gilovich, T. (2010), “The ball don’t lie: How inequity aversion can undermine performance”, *Journal of Experimental Social Psychology*, 46(6), 1148-1150.
International Football Association Board, 2023, *Laws of the Game*.
Mascarenhas, D. R., Collins, D., and Mortimer, P. (2002). “The art of reason versus the exactness of science in elite refereeing: Comments on Plessner and Betsch (2001)”. *Journal of Sport and Exercise Psychology*, 24(3), 328-333.
Moskowitz, T., and Wertheim, L. J. (2011). *Scorecasting: The hidden influences behind how sports are played and games are won*. Crown Archetype.
Nevill A, Balmer N., Williams M. (2002), “The influence of crowd noise and experience upon refereeing decisions in football”. *Psychology of Sport and Exercise*, 3:261–272.
Noecker, C. A., and Roback, P. (2012). “New insights on the tendency of NCAA basketball officials to even out foul calls”. *Journal of Quantitative Analysis in Sports*, 8(3).
Plessner, H. and Betsch, T. (2001), “Sequential Effects in Important Referee Decisions: The Case of Penalties in Soccer”. *Journal of Sport and Exercise Psychology*, pp. 254-259.
Reade, J. J., Schreyer, D., and Singleton, C. (2022). “Eliminating supportive crowds reduces referee bias”, *Economic Inquiry*, 60(3), 1416–1436.
Robberechts, P., Van Haaren, J., and Davis, J. (2019). “Who will win it? An in-game win probability model for football”, *arXiv preprint*:1906.05029, 4.
Schwarz, W. (2011). “Compensating tendencies in penalty kick decisions of referees in professional football: Evidence from the German Bundesliga 1963–2006”, *Journal of Sports Sciences*, 29(5), 441-447.
Scoppa, V. (2008). “Are subjective evaluations biased by social factors or connections? An econometric analysis of soccer referee decisions”. *Empirical Economics*, 1, 123-140.
Scoppa, V. (2021). “Social pressure in the stadiums: Do agents change behavior without crowd support?”, *Journal of Economic Psychology*, 82, 102344.
Sutter, M., and Kocher, M. G. (2004), “Favoritism of agents–the case of referees' home bias”, *Journal of Economic Psychology*, 25(4), 461-469.

Appendix

Table A1. Matches included in the Sample by League and Season

Season	England	France	Germany	Italy	Spain	Total
1213	294	364	306	379	380	1,723
1314	377	379	293	377	380	1,806
1415	377	378	306	380	380	1,821
1516	373	376	306	379	379	1,813
1617	380	379	305	380	380	1,824
1718	380	379	306	380	380	1,825
1819	380	380	306	380	380	1,826
1920	380	279	306	379	380	1,724
2021	379	380	306	380	375	1,820
2122	378	380	305	380	380	1,823
2223	373	378	306	380	375	1,812
2324	372	301	305	376	379	1,735
Total	4,443	4,353	3,656	4,552	4,548	21,552

Table A2. The Econometric Analysis using the Assigned Injury Time

	(1) Assigned Injury Time	(2) Assigned Injury Time	(3) Assigned Injury Time	(4) Assigned Injury Time	(5) Assigned Injury Time
Home Behind	0.390***	0.340***	0.327***	0.307***	0.338***
	(0.036)	(0.035)	(0.034)	(0.036)	(0.031)
Away Behind	0.211***	0.168***	0.159***	0.128***	0.169***
	(0.030)	(0.027)	(0.027)	(0.030)	(0.032)
# Substitutions		0.195***	0.193***	0.188***	0.104***
		(0.031)	(0.030)	(0.029)	(0.014)
# Injuries		0.432***	0.433***	0.430***	0.411***
		(0.034)	(0.033)	(0.032)	(0.024)
#VAR Reviews		0.970***	0.878***	0.863***	0.717***
		(0.084)	(0.077)	(0.076)	(0.064)
# Red Cards			0.237***	0.237***	0.295***
			(0.039)	(0.038)	(0.039)
# Yellow Cards			0.089***	0.087***	0.077***
			(0.015)	(0.015)	(0.010)
# Penalties			0.182***	0.119**	0.149***
			(0.046)	(0.041)	(0.042)
# Fouls			-0.047***	-0.048***	-0.015***
			(0.008)	(0.008)	(0.004)
# Shots				-0.020***	-0.017***
				(0.006)	(0.004)
# Goals				0.108**	0.095***
				(0.035)	(0.024)
Prob. Home Team Win				0.086	0.114
				(0.057)	(0.059)
Constant	4.089***	2.596***	2.875***	3.016***	2.563***
	(0.110)	(0.184)	(0.187)	(0.172)	(0.120)
Observations	10673	10673	10673	10669	10669
Adjusted R^2	0.013	0.211	0.235	0.243	0.392

Notes: The Table reports OLS estimates. The dependent variable is *Assigned Injury Time*. The variables *Home Behind* and *Away Behind* are evaluated at the end of regular time (90th minute). The number of events (substitutions, red cards, yellow cards, and so on) refer to the second half. Standard errors (reported in parentheses) are adjusted for heteroskedasticity and clustered at the league and season level. The symbols ***, ** and * indicate that coefficients are statistically significant at the 1%, 5%, and 10% levels, respectively.