**UNIVERSITA' DELLA CALABRIA**

Dipartimento di Economia, Statistica e Finanza "Giovanni Anania"

**Dottorato di Ricerca in**

**Scienze Economiche e Aziendali**

*Con il contributo di (Ente finanziatore)*

**Università della Calabria**

**CICLO: XXXVI**

**TITOLO TESI: The Economics of Soccer: Gender Strategies, Referee Fairness, and Home Advantage in Professional Soccer Leagues**

**Settore Scientifico Disciplinare: SECS-P/06**

**Coordinatore:** Ch.mo Prof. **Fabio Piluso**

Firma ________ FABIO PILUSO 25.03.2024 13:35:57 GMT+00:00

**Supervisore/Tutor**: Ch.mo Prof. **Vincenzo Scoppa**

Firma______________________

**Dottorando**: Dott. **Leo Morabito**

Firma ______________________

Università della Calabria

Dipartimento di Economia, Statistica e Finanza "*Giovanni Anania*"

**Dottorato di Ricerca in Scienze Economiche e Aziendali**

CICLO XXXVI

---

**Thesis**

# *The Economics of Soccer: Gender Strategies, Referee Fairness, and Home Advantage in Professional Soccer Leagues*

---


*Dottorando*

**Dott. MORABITO Leo**

*Supervisor*

**Prof. SCOPPA Vincenzo**

# Abstract

This thesis presents a multifaceted exploration into the realm of behavioral economics through the lens of professional soccer. The choice of soccer as a data source stems from the challenge of capturing authentic behavioral data in occupational contexts. In the pursuit of validating economic theories, real-world data becomes invaluable, and professional sports, especially soccer, offer a rich and accessible repository. Soccer, a globally revered sport, not only showcases athletes' skills but also serves as an intriguing subject due to the blend of monetary and non-monetary incentives driving the players. The thesis comprises three distinct studies, each shedding light on different behavioral aspects of professional soccer. Collectively, these studies offer profound insights into the behavioral dynamics in high-stakes, high-pressure environments. They underscore the complexity of human decision-making, influenced by various factors ranging from gender and risk tolerance to social pressures and fairness concerns. The findings not only enrich the academic discourse in sports economics and behavioral economics but also have practical implications for understanding and managing behaviors in professional sports.

# Contents


**List of Tables** **iii**

**1 Mixed-Strategy Equilibria and Gender Differences: the Soccer Penalty Kick game** **1**

1.1 Introduction . . . . . 2

1.2 The Soccer Penalty Kick game . . . . . 5

1.2.1 The rules of the game . . . . . 5

1.2.2 Mixed-strategy Nash equilibrium of the game . . . . . 7

1.3 Data and Test of the Assumptions . . . . . 8

1.3.1 Testing the Natural Side Assumption . . . . . 10

1.3.2 Testing the Identical Goalkeepers Assumption . . . . . 11

1.3.3 Testing the Simultaneous-moves Game Assumption . . . . . 12

1.4 Testing the Mixed Strategy Nash Equilibrium . . . . . 13

1.4.1 Estimation and Results . . . . . 15

1.4.2 Re-evaluating the Center option in Penalty Kicks . . . . . 18

1.5 Discussion and Concluding Remarks . . . . . 19

References . . . . . 21

**2 The Pursuit of Fairness in Ongoing Evaluations: Evidence from Referees in Soccer** **25**

2.1 Introduction . . . . . 26

2.2 Data and Descriptive Statistics . . . . . 28

2.3 Methodology . . . . . 29

2.4 Evidence of Even-out Bias of Referees . . . . . 31

2.4.1 Penalty Kicks . . . . . 31

2.4.2 Red Cards . . . . . 35

2.4.3 Yellow Cards . . . . . 38

2.4.4 Methodological Checks . . . . . 40

2.5 Discussion . . . . . 42

References . . . . . 43

**3 Home advantage and Gender Differences: Evidence from Major Women's European leagues** **45**

3.1 Introduction . . . . . 46

3.2 Data and Descriptive Statistics . . . . . 48

3.3 Home Advantage and Team Performance . . . . . 50

3.3.1 Points . . . . . 50

3.3.2 Goals . . . . . 51

3.4 Home Advantage and Referee's Gender . . . . . 52

3.4.1 Penalty Kicks . . . . . 53

3.4.2 Red Cards . . . . . 54

3.4.3 Yellow Cards . . . . . 55

3.5 Conclusion . . . . . 57

References . . . . . 58

# List of Tables


1.1 The Soccer Penalty Kick game: General Randomization . . . . . . . . . . . . . 5
1.2 The Soccer Penalty Kick game: Restricted Randomization . . . . . . . . . . . 6
1.3 Descriptive statistics . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9
1.4 Foot-types and Preferences of Kickers . . . . . . . . . . . . . . . . . . . . . 10
1.5 Testing the Assumption that Goalkeepers are Homogeneous . . . . . . . . . . 12
1.6 Testing The Assumption of Simultaneous-moves Game . . . . . . . . . . . . . 13
1.7 Regression Results - GR model . . . . . . . . . . . . . . . . . . . . . . . . . 16
1.8 3x3 game: Female players . . . . . . . . . . . . . . . . . . . . . . . . . . . . 16
1.9 3x3 game: Male players . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17
1.10 Regression Results - RR model where C = NS . . . . . . . . . . . . . . . . . 17
1.11 2x2 game: Female players . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17
1.12 2x2 game: Male players . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 18
1.13 Regression Results - RR model where C = OS . . . . . . . . . . . . . . . . . 19
1.14 2x2 game: Female players . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19
1.15 2x2 game: Male players . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19
A1 Foot-types and Preferences of Goalkeepers . . . . . . . . . . . . . . . . . . . 23
A2 Joint significance of *goalie-fixed effects* across different restrictions . . . . . . . 23

2.1 Descriptive Statistics . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 29
2.2 Observed and Predicted number of matches with 0, 1, 2, 3, and 4 Penalty Kicks 32
2.3 Transition matrix of the matches with exactly two Penalty Kicks . . . . . . . . 32
2.4 Even-out bias of Referees on awarding a Penalty Kick . . . . . . . . . . . . . 33
2.5 Observed and Predicted number of matches with 0, 1, 2, 3, and 4 red cards . 36
2.6 Transition matrix of the matches with exactly two Red Cards . . . . . . . . . 36
2.7 Even-out bias of Referees on issuing a Red Card . . . . . . . . . . . . . . . . 37
2.8 Even-out bias of Referees on issuing a Yellow Card . . . . . . . . . . . . . . . 40
2.9 Robustness check on the Even-out bias . . . . . . . . . . . . . . . . . . . . . 41
B1 Number of Games per season across Top 5 European Leagues . . . . . . . . . 44

3.1 Data and Attendance Composition . . . . . . . . . . . . . . . . . . . . . . . 49
3.2 Descriptive Statistics . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 50
3.3 Home Advantage: Points . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 52
3.4 Home Advantage: Goals . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 53
3.5 Referee's gender and Home Advantage: Penalty Kicks . . . . . . . . . . . . . 54
3.6 Referee's gender and Home Advantage: Red Cards . . . . . . . . . . . . . . . 55
3.7 Referee's gender and Home Advantage: Yellow Cards . . . . . . . . . . . . . 56

# Chapter 1

# Mixed-Strategy Equilibria and Gender Differences: the Soccer Penalty Kick game

## Abstract

This research employs a natural experiment to explore gender differences using mixed strategies during penalty kicks, extending the game model introduced by Chiappori et al. (2002). By analyzing data from professional male and female soccer players in parallel soccer competitions, this study uncovers a significant divergence in strategic approaches. Female players show notable avoidance of the Center (C) option, leading to a rejection of the hypothesis that they play the 3x3 game, aligning more with a 2x2 gameplay model; At the same time, men seem to follow the theory remarkably. The divergent strategic choices observed in penalty kicks across genders can be partially explained by the *Action bias* detailed by Bar-Eli et al. (2007). Despite these variations in strategy, the statistical probability of scoring a penalty kick remains consistent across genders. This research enriches the literature by highlighting the influence of gender on strategic choices in high-pressure contexts. It demonstrates that behavioral biases can shape the decision-making process without altering performance outcomes.

**JEL Classification:** C72, D81, D91, J16
**Keywords:** Mixed-strategy Nash equilibrium; Penalty kicks game; Gender differences; Choice behavior; Sport psychology; Behavioral economics.

## 1.1 Introduction

In recent decades, game theory has become increasingly vital for understanding how economic agents make decisions, especially in scenarios marked by strategic interdependence. Since Nash's seminal work in 1951, the concept of non-cooperative games—where players act independently without forming coalitions or engaging in collaboration—has been central to game theory. In these settings, the Mixed-Strategy Nash Equilibrium (MSNE) plays a crucial role. In mixed-strategy equilibria, players randomize across several actions, i.e., assign a certain probability to every possible pure strategy. Each player's mixed strategy is optimal given the equilibrium behavior of the other players. In other words, no player has incentives to deviate from a mixed strategy, as long as the other players choose their equilibrium strategies.

Laboratory experiments have been essential for the controlled testing of game theory principles. However, their effectiveness in exploring mixed-strategy Nash equilibrium scenarios has been mixed, yielding ambiguous results.

For example, O'Neill (1987) tests the MSNE theory in a laboratory experiment on students playing a repeated 2 x 2 zero-sum game finding conclusions supporting the actual use of mixed strategies[1]. However, these findings were later contested by Brown and Rosenthal (1990). Shachat (2002) reported a series of experiments aimed at discriminating between possible sources of failure of the unique mixed equilibrium of O'Neill's game, concluding that while the MSNE hypothesis fails to fully rationalize the behavior of individuals, there is still no more robust alternative theory. Rapoport and Budescu (1992) noted that individuals often struggle to generate an uncorrelated series of actions, questioning the practical application of MSNE.

The poor aptitude of individuals to follow the MSNE predictions during an experiment can be traced back to the subjects' lack of experience in playing and being consistent with a randomization strategy of the actions. Student subjects participating in lab experiments no doubt understand the rules, but they have neither the experience, the time, nor the incentive to learn to play well (Gauriot et al.; 2023).

This is why studies in natural contexts have often yielded more consistent results regarding the use of mixed strategies. In real-life scenarios, such as tax control[2], diplomacy, or warfare, unpredictability is fundamental. These are arenas where strategic competitive interactions demand that players' actions remain unpredictable to their opponents. By looking at how experts play games, we can bypass the complexities of learning Azar and Bar-Eli (2011).

The best evidence supporting mixed strategies in pure-conflict games comes from professional sports, where players, in addition to having great experience, show their intrinsic desire to win, reinforced by large monetary and non-monetary incentives.

The first application of mixed strategies in sports scenarios was carried out by Walker and Wooders (2001) who tested the use of mixed strategies in tennis by studying the direction of

[1] A simultaneous game between two players and with monetary winnings which consists of choosing a card from a set of 4 cards.

[2] Tax control agencies may randomize their choice of whom to audit taxes, and taxpayers may decide to maximize their output by randomizing their decision to truthfully report their income Azar and Bar-Eli (2011)

serving and the set win rates for 15 matches of tennis played at the Wimbledon tournament. They concluded that tennis players, in deciding on which side of the receiver to serve the ball, follow the predictions of the MSNE, but do not satisfy the condition of serial independence. More convincing results were found by Chiappori et al. (2002), Palacios-Huerta (2003), Coloma (2007) and Azar and Bar-Eli (2011) in the study of the Soccer Penalty Kick game involving male soccer players.

In detail, Chiappori et al. (2002) analyzed a sample of 459 penalty kicks, observing the actions of both players – the kicker and the goalkeeper. They modeled penalty kicks as a 3x3 game and examined the scoring probabilities for each action, Right, Center, and Left, along with other implications for aggregated data. This included the assumptions that the players play simultaneously and that goalkeepers can be considered homogeneous in their approach to the game. In such a simultaneous-moves game, it is highly plausible that players conform to mixed strategies theory, as they should randomize their actions to maximize their outcome and to prevent the opponent from a winning strategy. Their findings are consistent with these assumptions.

Conversely, Palacios-Huerta (2003) approached it as a 2x2 game, where players can choose between their preferred side or Natural Side (NS) and their non-preferred side or Opposite Side (OS), assuming that a central option can be considered part of the NS. In this case, he tested the hypothesis of equal scoring probabilities between the two actions on individual players instead of on aggregate data and if the players can generate random sequences of choices, finding very favorable results. Similarly, Coloma (2007), using the same data of Palacios-Huerta (2003), developed a method to directly test the predictions of mixed strategies using a system of linear probability models, finding highly consistent results. Finally, Azar and Bar-Eli (2011) explored different alternative decision-making processes that players could use during the Penalty Kick game, finding the MSNE the most favorable theory.

Professional football players likely understand the importance of generating random patterns to maximize their payoffs during penalty kicks. It is no secret that professional goalkeepers often study the penalties previously taken by their opponents to make inferences about the next[3].

However, most existing studies on mixed strategies have utilized natural experimental contexts featuring male athletes, leaving a gap in understanding gender differences in strategic gameplay. Recent studies have further explored gender differences in MSNE adherence. Notably, Gauriot et al. (2023) analyzed a dataset of 500,000 tennis serves, finding that while men's plays align consistently with mixed strategies, women's adherence to the theory is less pronounced. Complementing this, Jeon and Ong (2020) achieved similar conclusions in a laboratory experiment regarding auction bids, suggesting that the observed gender differences in competitive contexts could be attributed to variations in risk perception. This aligns with Goeree et al. (2003) proposition that incorporating risk aversion into game strategy estimations can reconcile discrepancies between Nash equilibrium predictions and experimental data. For instance, when players have different attitudes toward risk or other

[3] A famous attempt to use past actions to guess how the game will be played is that of the English goalkeeper Jordan Pickford. During the European final England - Italy, he consulted the list of Italian penalty kickers with the related preferred shooting directions.

preference characteristics, their utility functions may be different, as well as the equilibrium predictions.

The literature on gender-based differences in risk attitudes is extensive and generally concludes that women exhibit greater risk aversion compared to men in decision-making under uncertainty, as evidenced in studies by Eckel and Grossman (2008), Charness and Gneezy (2012), and Gneezy et al. (2003).

Similarly, Paserman (2011) examined gender approaches in high-stakes tennis matches. His findings indicated that women are more likely to commit errors at critical game junctures, possibly due to adopting more conservative strategies in crucial moments.

These studies imply that gender differences in risk preferences and competitive strategies are significant, especially in high-pressure sports contexts. They suggest that understanding these differences is crucial for a comprehensive grasp of decision-making dynamics in competitive environments.

Given these insights, the primary objective of this research is to investigate whether similar gender differences manifest during the play of the Penalty Kick game in soccer, where pressure and risk are very relevant for the outcome. In fact, in such a context, some strategies are perceived as relatively safe and represent the norm; for example, as we will see later on, kicking toward one of the sides could represent a relatively safe strategy because it involves a fairly high payoff on average whether or not the goalkeeper anticipates the action. In case of failure, the kicker is not blamed by the supporters, but they often congratulate the opponent. Instead, kicking toward the center may be perceived as a risky action: almost in all cases if the goalkeepers decide to stay in the center of the goal, they manage to save the penalty. In this case, supporters and teammates probably blame the risky choice of the kickers. The same can be said from the goalkeeper's point of view; staying in the center implies high utility if the choice of the kicker coincides, but very low if the kicker kicks toward one of the two sides because of the inaction.

Bar-Eli et al. (2007), in their analysis of a 3x3 penalty kick game, identified an *Action bias* in goalkeepers, who are more inclined to jump rather than stay in the center during penalty kicks. This inclination is tied to *Norm Theory* proposed by Kahneman and Miller (1986). This theory suggests that individuals anticipate greater regret from negative outcomes resulting from action versus inaction. In penalty kicks, action (jumping or kicking to a side) is considered the norm; a bias towards action indicates a psychological preference for making a visible effort, driven by the belief that active efforts to stop the penalty will face less criticism and regret, even if unsuccessful. It helps to highlight the thought processes of players in high-pressure situations, where taking action is favored over inaction. Such behavior challenges the mixed-strategy theory, which assumes players are indifferent across potential strategies.

This research paper is structured as follows: Section 2 describes the game-theoretical model for penalty kicks, outlining the underlying assumptions and hypotheses. Section 3 presents basic statistics and the first tests of the assumptions of the Penalty Kick game. Section 4 details the empirical test of the Mixed Strategy Equilibrium, and Section 5 offers a discussion of the findings.

## 1.2 The Soccer Penalty Kick game

A penalty kick consists of kicking the ball from the penalty spot directly into the opponent's goal which is protected only by the opposing goalkeeper. There are two occasions when the penalty kick is used during a soccer match: the first is when one of the fouls punishable under the Fédération Internationale de Football Association (FIFA) rules is committed inside the penalty area; the second occasion, called shootouts, occurs only in knockout matches or finals when the match ends in a draw, and it must be decided which team passes the round or wins the competition.

### 1.2.1 The rules of the game

From a game theory perspective, the Penalty kicks game can be considered a *simultaneous, non-cooperative* game without a *pure-strategy equilibrium*. It is a simultaneous moves game because the players must decide their actions before the kicker kicks the ball. This is because the average travel time of the ball from the penalty spot to the goal (11 meters) is around 0.2 and 0.3 seconds, with a speed that can reach 130 km/h (Chiappori et al.; 2002). Generally, the goalkeeper cannot wait and see the direction of the kick but has to decide the jumping strategy before the penalty is shot. We will delve deep into this assumption later when we will empirically test it. It is a non-cooperative game as the kicker's win corresponds to the goalkeeper's loss and vice versa. Finally, there are no pure strategies: if a player consistently chooses the same strategy (e.g., always shooting or jumping to the right), their opponent would exploit it and win the game.

In this game, the action space $S$ consists of the pure strategies available to both players. From the kicker's perspective, the options are to kick to the Left (L), Center (C), or Right (R). Similarly, the goalkeeper's choices are to dive to the kicker's Left (L), stay in the Center (C), or to the kicker's Right (R).

#### Assumption of kicker's Natural Side

According to Chiappori et al. (2002) and Palacios-Huerta (2003), kickers' preferences for a kicking side are influenced by their dominant foot. Typically, left-footed players prefer to kick toward their right, and vice versa. This assumption, rooted in an observable variable, allows for the empirical verification of the game's symmetry across different foot types. If the assumption holds, observations can be aggregated regardless of the foot type of the player to conceptualize a 3x3 game, also called General Randomization (GR). In this model, the available strategies are defined as the kicker's Natural Side (NS), the Center (C), and the Opposite Side (OS), as illustrated in Table 1.1.

| | Goalkeeper | | |
|---|---|---|---|
| Kicker | **NS** ($q_{NS}$) | **C** ($q_C$) | **OS** ($q_{OS}$) |
| **NS** ($p_{NS}$) | $\theta_{NS}$ | $\pi_{NS}$ | $\pi_{NS}$ |
| **C** ($p_C$) | $\mu$ | 0 | $\mu$ |
| **OS** ($p_{OS}$) | $\pi_{OS}$ | $\pi_{OS}$ | $\theta_{OS}$ |

**Table 1.1:** *The Soccer Penalty Kick game: General Randomization*

The six values in the General Randomization matrix represent the payoffs of the kicker given by the different scoring probabilities: $\theta_{NS}(\theta_{OS})$ is the probability for the kicker to score if both the kicker and the goalkeeper choose the Natural side (Opposite side); $\pi_{NS}(\pi_{OS})$ is the probability to score if the kicker kicks to the Natural side (Opposite side) and the goalkeeper jumps to the Center or the Opposite side (Natural side); $\mu$ is the probability of scoring if the kicker kicks to the Center and the goalkeeper jumps to one of the sides. We assume that this payoff matrix, in which the goalkeeper's payoff is simply 1 minus the probability of scoring the penalty, is known by both players involved. We also assume that $\pi > \mu > \theta > 0$. $p_S$ and $q_S$, with $S = \{NS, C, OS\}$, represent the probabilities that the kicker and the goalkeeper choose, respectively, NS, C, and OS.

In literature, the use of the Center (C) strategy is debated. Palacios-Huerta (2003) notes that its inclusion in the model does not significantly alter the game's outcome. Professional players, as mentioned by Palacios-Huerta, perceive kicking to the center and their natural side as similarly challenging, primarily because they use the inside of their foot for better control. Despite the potential for a third option, the aggregate analysis by Chiappori et al. (2002) concludes that the presence of the center as a pure strategy does not raise concerns. These findings support the argument that a penalty kick can be aptly modeled as a 2x2 game, the Restricted Randomization (RR), as shown in Table 1.2

| | | Goalkeeper | |
|---|---|---|---|
| | | **NS** ($q_{NS}$) | **OS** ($q_{OS}$) |
| Kicker | **NS** ($p_{NS}$) | $\theta_{NS}$ | $\pi_{NS}$ |
| | **OS** ($p_{OS}$) | $\pi_{OS}$ | $\theta_{OS}$ |

**Table 1.2:** *The Soccer Penalty Kick game: Restricted Randomization*

To discern whether players are engaging in a 2x2 or a 3x3 game, Chiappori et al. (2002) formulates the a General Randomization proposition:

<u>GR Proposition</u>: *If it holds that:*

$$\mu > \frac{\pi_{NS} \cdot \pi_{OS} - \theta_{NS} \cdot \theta_{OS}}{\pi_{NS} + \pi_{OS} - \theta_{NS} - \theta_{OS}} \tag{1.1}$$

*then both players randomize over {NS, C, OS}, the General Randomization (3x3). Otherwise, both players randomize over {NS, OS}, the Restricted Randomization (2x2).*

The rationale behind this proposition is quite straightforward. If the success rate (or outcome) of kicking to the center, $\mu$, is sufficiently low, the kicker will be inclined to choose this option with a very minimal probability. Consequently, the goalkeeper, being aware of the kicker's tendency, will also rarely decide to choose the Center. The game would be reduced to a 2x2 game with just two pure strategies, NS and OS, for each player. Conversely, if the kicker perceives a substantially higher payoff from choosing C while the goalkeeper is biased towards the sides, the kicker will opt for C with a notable probability. In equilibrium, each player's strategy should allocate the probabilities in such a way that the other player is indifferent among all the options that are *actually used* during the game (Dixit et al.; 2020).

**Assumption of Identical Goalkeepers.**

In real-life scenarios, match-specific probabilities are not observable because rarely a kicker faces the same goalkeeper more than a few times during the season. For this reason, to empirically test the application of mixed strategies, we have to take account of the heterogeneity. We do this by testing the assumption that the goalkeepers play the game homogeneously. That is:

*For any match between a kicker i and a goalkeeper j, the parameters $\pi_{NS}$, $\pi_{OS}$, $\theta_{NS}$, $\theta_{OS}$ and $\mu$ do not depend on j.*

This assumption, although very strong, is not entirely implausible. It should not be confused with the hypothesis that goalkeepers are equally performing in their role, but that when faced with penalties they generally act in the same way, resulting in the same probability of saving a penalty.

If this assumption holds, there are some empirical consequences:

1. The kicker's strategy does not depend on the goalkeeper;
2. The goalkeeper's strategy is identical for all goalkeepers;
3. The scoring probability is the same whether the kicker kicks NS, C, or OS, irrespective of the goalkeeper.

This assumption is crucial because it allows us to test the predictions of the model by considering the game kicker-specific. That is, while kickers differ from each other in abilities or expertise, goalkeepers can be considered identical for each observation.

### 1.2.2 Mixed-strategy Nash equilibrium of the game

In a mixed strategy Nash equilibrium each player must be indifferent among the available options, i.e. each action should yield the same expected payoff. It implies that for both the kicker (K) and the goalkeeper (G) in the equilibrium:

$$\mathbb{E}[NS_K] = \mathbb{E}[C_K] = \mathbb{E}[OS_K] \tag{1.2}$$

and

$$\mathbb{E}[NS_G] = \mathbb{E}[C_G] = \mathbb{E}[OS_G] \tag{1.3}$$

Using the payoffs matrix of the General Randomization game (Tab. 1.1), in equilibrium the probabilities that the kicker chooses NS ($p_{NS}$), OS ($p_{OS}$) and C ($p_C$) are, respectively:

$$p_{NS} = \frac{\mu \cdot (\pi_{OS} - \theta_{OS})}{(\pi_{OS} - \theta_{OS}) \cdot (\pi_{NS} - \theta_{NS}) + \mu \cdot (\pi_{NS} + \pi_{OS} - \theta_{NS} - \theta_{OS})}, \tag{1.4}$$

$$p_{C} = \frac{(\pi_{OS} - \theta_{OS}) \cdot (\pi_{NS} - \theta_{NS})}{(\pi_{OS} - \theta_{OS}) \cdot (\pi_{NS} - \theta_{NS}) + \mu \cdot (\pi_{NS} + \pi_{OS} - \theta_{NS} - \theta_{OS})}, \tag{1.5}$$

$$p_{OS} = \frac{\mu \cdot (\pi_{NS} - \theta_{NS})}{(\pi_{OS} - \theta_{OS}) \cdot (\pi_{NS} - \theta_{NS}) + \mu \cdot (\pi_{NS} + \pi_{OS} - \theta_{NS} - \theta_{OS})} \tag{1.6}$$

whereas the probabilities that the goalkeeper chooses NS ($q_{NS}$), OS ($q_{OS}$) and C ($q_C$) are:

$$q_{NS} = \frac{\pi_{NS} \cdot (\pi_{OS} - \theta_{OS}) + \mu \cdot (\pi_{NS} - \pi_{OS})}{(\pi_{OS} - \theta_{OS}) \cdot (\pi_{NS} - \theta_{NS}) + \mu \cdot (\pi_{NS} + \pi_{OS} - \theta_{NS} - \theta_{OS})}, \tag{1.7}$$

$$q_C = \frac{\mu \cdot (\pi_{NS} + \pi_{OS} - \theta_{NS} - \theta_{OS}) + \theta_{NS} \cdot \theta_{OS} - \pi_{NS} \cdot \pi_{OS}}{(\pi_{OS} - \theta_{OS}) \cdot (\pi_{NS} - \theta_{NS}) + \mu \cdot (\pi_{NS} + \pi_{OS} - \theta_{NS} - \theta_{OS})}, \tag{1.8}$$

$$q_{OS} = \frac{\pi_{OS} \cdot (\pi_{NS} - \theta_{NS}) - \mu \cdot (\pi_{NS} - \pi_{OS})}{(\pi_{OS} - \theta_{OS}) \cdot (\pi_{NS} - \theta_{NS}) + \mu \cdot (\pi_{NS} + \pi_{OS} - \theta_{NS} - \theta_{OS})} \tag{1.9}$$

Instead, if the *GR Proposition* [Eq. 1.1] does not hold, the players consider the Restricted Randomization game in which the Center is not an option; then the probabilities of the kicker choosing NS ($p_{NS}$) and OS ($p_{OS}$) are, respectively:

$$p_{NS} = \frac{\pi_{OS} - \theta_{OS}}{\pi_{NS} + \pi_{OS} - \theta_{NS} - \theta_{OS}} \tag{1.10}$$

and

$$p_{OS} = 1 - p_{NS} = \frac{\pi_{NS} - \theta_{NS}}{\pi_{NS} + \pi_{OS} - \theta_{NS} - \theta_{OS}} \tag{1.11}$$

while the probabilities that the goalkeeper chooses NS ($q_{NS}$) and OS ($q_{OS}$) are, respectively:

$$q_{NS} = \frac{\pi_{NS} - \theta_{OS}}{\pi_{NS} + \pi_{OS} - \theta_{NS} - \theta_{OS}} \tag{1.12}$$

and

$$q_{OS} = 1 - q_{NS} = \frac{\pi_{OS} - \theta_{NS}}{\pi_{NS} + \pi_{OS} - \theta_{NS} - \theta_{OS}} \tag{1.13}$$

Moreover, an integral hypothesis is the randomization of players' choices. This hypothesis posits that each player's decision-making process is characterized by serial independence when playing the game. This independence is crucial for ensuring the unpredictability and fairness of the game, as it precludes players from developing deterministic strategies based on historical patterns.

## 1.3 Data and Test of the Assumptions

Our dataset comprises 4297 penalty kicks collected from professional games across various competitions and countries, spanning the years 2016 to 2022. To effectively analyze gender differences, we gathered data from equivalent top leagues and competitions for both male and female players[4]. Specifically, the dataset includes 3386 penalty kicks from male players and 911 from female players. The huge difference between the two samples reflects the fewer games played in female competitions.

[4] In detail, the country leagues we considered are the Division 1 Féminine and Ligue 1 in France, Bundesliga and Frauen-Bundesliga in Germany, Premier League and Women's Super League in England, and Serie A and Serie A Femminile in Italy. In addition to these, observations from the Champions League, World Cup, and European Championship are also included for both female and male players.

We included both game-specific and player-specific variables. Game-specific variables include the season in which the match was played, the location from the kicker's perspective (home or away), the minute of the penalty kick, the score before the penalty, whether it was taken during shootouts, and the outcome (scored or missed). Player-specific variables consist of the identity of the involved players, their roles, their teams, and the kicker's preferred foot.

All this information was meticulously gathered from various football-related websites, such as *transfermarkt.com* and *fbref.com*. Additionally, the direction of the penalty kick and the goalkeeper's jump direction were documented through the analysis of video highlights of soccer matches. Table 1.3 shows the frequencies and the scoring rate relative to the most important variables collected.

**Table 1.3:** *Descriptive statistics*

| | Females | | | Males | | |
|---|---|---|---|---|---|---|
| *Variables* | Obs. | Frq. (%) | **Scoring Rate** | Obs. | Frq. (%) | **Scoring Rate** |
| *Competition* | | | | | | |
| France | 161 | 17.67 | **0.81** | 637 | 18.81 | **0.79** |
| Germany | 68 | 7.46 | **0.75** | 194 | 5.73 | **0.80** |
| England | 93 | 10.21 | **0.72** | 501 | 14.80 | **0.79** |
| Italy | 154 | 16.90 | **0.84** | 726 | 21.44 | **0.79** |
| USA | 157 | 17.23 | **0.69** | 791 | 23.36 | **0.78** |
| UCL | 90 | 9.88 | **0.79** | 316 | 9.33 | **0.77** |
| National | 188 | 20.64 | **0.71** | 221 | 6.53 | **0.71** |
| *Circumstances* | | | | | | |
| In-game | 786 | 86.28 | **0.76** | 3093 | 91.35 | **0.78** |
| Shootout | 125 | 13.72 | **0.73** | 293 | 8.65 | **0.75** |
| *Kicker's location* | | | | | | |
| Home | 482 | 52.91 | **0.78** | 1916 | 56.59 | **0.79** |
| Away | 429 | 47.09 | **0.73** | 1470 | 43.41 | **0.77** |
| *Kicker's role* | | | | | | |
| Defender | 114 | 12.51 | **0.77** | 185 | 5.46 | **0.78** |
| Midfielder | 315 | 34.58 | **0.77** | 1161 | 34.29 | **0.79** |
| Forward | 482 | 52.91 | **0.74** | 2040 | 60.25 | **0.78** |
| *Foot-type* | | | | | | |
| Left | 141 | 15.48 | **0.79** | 739 | 21.83 | **0.80** |
| Right | 770 | 84.52 | **0.75** | 2647 | 78.17 | **0.78** |
| Total | 911 | | **0.76** | 3386 | | **0.78** |

The observations exhibit comparable representation frequencies among female and male competitions. A notable exception is observed in penalties within national leagues, constituting over 20% of the penalties in female leagues compared to a mere 6.53% in the male sample of observations. A significant disparity is also evident in the scoring rates. Male scoring rates demonstrate greater uniformity across various competitions, whereas those for females exhibit notable variability ($F$=2.515, $p<0.05$), with the penalty kick taken during national games being the worst in terms of kicker's output. Conversely, no statistically significant differentiation is discernible between penalties kicked during regular match play and shootouts for either gender.

Similarly, there is no "home advantage" since for both genders the scoring rates are not statistically different. Both female and male observations display a parallel distribution of players by role, with forward players predominantly taking penalty shots, aligning with conventional expectations. However, the role of the player does not yield a statistically significant impact on penalty success rates. Moreover, in line with the real-life distribution of foot-type players, the sample indicates a higher incidence of penalties taken by right-footed players; yet, for both genders, the preferred foot does not statistically influence the scoring rate. Lastly, when comparing the scoring rates between genders, it is observed that women appear to have a lower success rate from the penalty spot. However, this difference is not statistically significant (t=1.53, p=0.125).

Having detailed our dataset, we now proceed to the empirical testing of the assumptions outlined in the previous section.

### 1.3.1 Testing the Natural Side Assumption

The first assumption we test concerns the preferences of kickers in striking the ball. As previously noted in the literature on penalty kicks, male professional players tend to cross their shots. To verify whether similar preferences exist among female professional players, statistically examine if shot direction correlates with the player's foot type using a chi-square test. The choices the kickers make are presented in Table 1.4.

Observing the choice percentages of the different foot types, it can be inferred that for both genders, there is a preference to cross the shot. Indeed, the table shows that 58.9% of penalty kicks made by left-footed female players went to the right, while right-footed female players shot to the left 57.7% of the time. This pattern of shooting away from their dominant foot side supports the 'Natural Side' assumption.

**Table 1.4:** *Foot-types and Preferences of Kickers*

| | Foot-type | Shot side | | | Total |
|---|---|---|---|---|---|
| | | **Left** | **Center** | **Right** | |
| **Females** | Left-footed | 52<br>*36.9%* | 6<br>*4.3%* | 83<br>*58.9%* | 141<br>*100 %* |
| | Right-footed | 444<br>*57.7%* | 58<br>*7.5 %* | 268<br>*34.8%* | 770<br>*100 %* |
| | Total | 496<br>*54.4%* | 64<br>*7 %* | 351<br>*38.5%* | 911<br>*100 %* |
| **Males** | Left-footed | 268<br>*36.3%* | 122<br>*16.5%* | 349<br>*47.2%* | 739<br>*100 %* |
| | Right-footed | 1276<br>*48.2%* | 436<br>*16.5 %* | 935<br>*35.3%* | 2647<br>*100 %* |
| | Total | 1544<br>*45.6%* | 558<br>*16.5 %* | 1284<br>*37.9%* | 3386<br>*100 %* |

*Notes*: $\chi^2_{\text{female}}(2) = 29.182, p = 0.000$, $\chi^2_{\text{male}}(2) = 39.648, p = 0.000$.

The results of the chi-square test for both genders ($\chi^2_{\text{female}} = 29.182$; $\chi^2_{\text{male}} = 39.648$) reveal a significant association between the kicker's foot type and their chosen shot direction. Similarly, goalkeepers make their choices by observing the kicker's foot type. They tend to choose the kicker's preferred side more frequently ($\chi^2_{\text{female}} = 20.637$, $p = 0.000$ and

$\chi^2_{\text{male}} = 120.015$, $p = 0.000$). The frequency table for goalkeepers' choices is provided in the Appendix (Tab. A1). Therefore, goalkeepers across both genders are aware of this preference; the concept of a 'Natural Side' can be applied to both genders.

However, the data also reveals a gender-based difference in the inclination to shoot towards the center, with men showing a greater preference for this choice. This difference is statistically significant (t=2.4, p<0.01). As highlighted in the introduction and supported by Bar-Eli et al. (2007), the observed reluctance of players to opt for the center in penalty shootouts may result from behavior influenced by risk aversion. This tendency not only reflects the players' strategic choices but also leads to significant outcomes in the application of mixed strategies in the Soccer Penalty Kick game.

### 1.3.2 Testing the Identical Goalkeepers Assumption

To evaluate whether the kicker's behavior is consistent with mixed strategy equilibrium play, it is crucial to consider the observed sequence of a kicker's actions as stemming from repeated play of the same game. This approach is valid if goalkeepers are not heterogeneous in playing the Penalty Kick game and the kicker acts like the goalkeepers are identical. By following Chiappori et al. (2002) and Dohmen and Sonnabend (2018) we test this assumption by running three linear probability models of the form:

$$Y_i = \beta_0 + \mathbf{X}_i\beta + \alpha_{\text{K}} + \gamma_{\text{G}} + \epsilon_i \tag{1.14}$$

The outcome variable $Y_i$ assumes three different variables: whether a kick is successful, whether the kicker kicks to his natural side (NS), and whether the goalkeeper jumps to the kicker's natural side (NS). $\boldsymbol{X}_i$ is a set of explanatory variables controlling for match and player characteristics such as home advantage, the score difference before the penalty, game quarter, kicker's role, and competition. $\alpha_K$ and $\gamma_G$ represent fixed effects for kickers and goalkeepers, respectively. The *null hypothesis*, which posits that all goalkeepers are identical, corresponds to the goalkeeper-fixed effects being jointly equal to zero. We employ an F-test to test this hypothesis. The analysis includes goalkeepers with at least four penalty kick observations. Results for both genders are presented in Table 1.5.

The F-statistics and their corresponding p-values for both genders (female: p=0.297, 0.974, 0.503; male: p = 0.406, 0.808, 0.605) in all regression models indicate that goalkeeper fixed effects are jointly insignificant. Therefore, we cannot reject the hypothesis that both female and male goalkeepers are homogeneous when playing the penalty kick game.

Furthermore, we performed additional regressions under varying restrictions on the minimum number of penalty kicks faced by individual goalkeepers, detailed in Table A2 in the Appendix. These tests reaffirm the initial findings, as the hypothesis of identical goalkeepers holds under most of the subsamples.[5]

[5] An exception is noted for female goalkeepers with fewer than three observations in the sample, where the hypothesis is rejected.

**Table 1.5:** *Testing the Assumption that Goalkeepers are Homogeneous*

| | | *Dependent variable:* | | |
|---|---|---|---|---|
| | | Kick Saved | Kicker shoots NS | Goalie jumps NS |
| **Females** | $F$ statistic: joint significance of goalie-fixed effects | 1.168 [$p$ = 0.297] | 0.692 [$p$ = 0.974] | 0.992 [$p$ = 0.503] |
| | $R^2$ | 0.637 | 0.574 | 0.578 |
| | Obs. | 694 | 694 | 694 |
| **Males** | $F$ statistic: joint significance of goalie-fixed effects | 1.020 [$p$ = 0.406] | 0.914 [$p$ = 0.808] | 0.971 [$p$ = 0.605] |
| | $R^2$ | 0.411 | 0.366 | 0.374 |
| | Obs. | 3105 | 3105 | 3105 |
| | Kicker-fixed effects? | yes | yes | yes |
| | Goalie-fixed effects? | yes | yes | yes |

*Notes*: The sample is limited to the goalkeepers with 4 or more penalty kicks in the dataset. The degrees of freedom of the *F test* are equal to 79, 293 for women and to 232, 1992 for men. If a goalkeeper is homogeneous, the *F test* should not reject the *null hypothesis* that all goalie-fixed effects are equal.

### 1.3.3 Testing the Simultaneous-moves Game Assumption

The essence of MSNE lies in its ability to capture strategic uncertainty in situations where players cannot observe others' actions before making a decision. For this reason, it is fundamental to test that players play the game simultaneously. Sequential play (i.e., one player acting after observing the opponent's action) would nullify mixed strategies, leading to predictable pure strategies. Moreover, such sequential action could be considered implausible in penalty kicks due to the rules [6] and the ball's fast travel time from the penalty spot to the goal, which leaves little time for the goalkeeper to react after the kicker's shot is made.

Prior research (Chiappori et al. (2002); Palacios-Huerta (2003); Dohmen and Sonnabend (2018)) has extensively validated this hypothesis for male professional players. We now extend this examination to both genders. To empirically assess the simultaneous-moves play, we use the following regression models:

$$NS_i^P = X_i\alpha + \beta NS_i^O + \gamma \overline{NS}_i^P + \sigma \overline{NS}_i^O + \epsilon_i \tag{1.15}$$

where $NS_i^P$ is a dummy variable indicating if the player (either the kicker or the goalkeeper) opts for the Natural Side in observation $i$. Similarly, $NS_i^O$ represents the opponent's (goalkeeper or kicker) action. The variables $\overline{NS}_i^P$ and $\overline{NS}_i^O$ denote the proportion of past actions to the Natural Side by the player and the opponent, respectively. The vector $\mathbf{X}$ includes control variables for match and player characteristics. In the context of simultaneous moves, the action chosen by one player on a penalty kick $i$, conditional on both players' history, should not be predictive of the opponent's action on the same penalty kick. This is equivalent to $\beta = 0$.

Table 1.6 presents the results from a series of linear probability models. Columns 2 and 6 of the table restrict the analysis to kickers who have taken at least four penalty kicks, whereas

[6] Kickers cannot feign a kick after completing the run-up. This strategy is prohibited by the rules of the penalty kick (see IFAB 2021-22, Law 14).

**Table 1.6:** *Testing The Assumption of Simultaneous-moves Game*

| | *Females* | | | | *Males* | | | |
|---|---|---|---|---|---|---|---|---|
| *Dependent variable:* | KNS | | GNS | | KNS | | GNS | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Constant | 0.719** | −0.262 | −0.051 | 0.150 | 0.363 | 0.206 | 0.233 | 0.458* |
| | (0.269) | (0.392) | (0.266) | (0.378) | (0.188) | (0.300) | (0.186) | (0.232) |
| Goalie jumps NS | 0.027 | 0.028 | | | −0.019 | −0.040 | | |
| | (0.050) | (0.071) | | | (0.022) | (0.027) | | |
| Kicker shoots NS | | | 0.026 | −0.020 | | | −0.019 | −0.027 |
| | | | (0.048) | (0.061) | | | (0.021) | (0.023) |
| Kicker's % of shots to NS, prior to this kick | 0.050 | 0.164 | 0.266*** | 0.293** | 0.099** | 0.257*** | 0.233*** | 0.230*** |
| | (0.070) | (0.171) | (0.067) | (0.088) | (0.035) | (0.062) | (0.034) | (0.038) |
| Goalie's % of jumps to NS, prior to this kick | −0.110 | −0.129 | −0.008 | 0.134 | 0.022 | −0.028 | 0.047 | 0.056 |
| | (0.077) | (0.111) | (0.075) | (0.144) | (0.044) | (0.057) | (0.044) | (0.064) |
| Kickers with 4+ obs? | no | yes | no | no | no | yes | no | no |
| Goalkeepers with 4+ obs? | no | no | no | yes | no | no | no | yes |
| Observations | 443 | 226 | 443 | 285 | 2,179 | 1,442 | 2,179 | 1,790 |
| $R^2$ | 0.066 | 0.124 | 0.091 | 0.094 | 0.014 | 0.026 | 0.029 | 0.029 |
| Adjusted $R^2$ | 0.010 | 0.020 | 0.036 | 0.007 | 0.003 | 0.010 | 0.018 | 0.015 |

*Notes:* The table reports OLS estimates. Standard errors are in parentheses. In columns 2 and 6, the samples are restricted to kickers who took more than four kicks. In columns 4 and 8, the samples are restricted to goalkeepers who were involved in more than four penalty kick situations. *p<0.05; **p<0.01; ***p<0.001

columns 4 and 8 show the results with the same sample limitation for the goalkeepers. The findings presented in Table 1.6 support the hypothesis that both players make their moves simultaneously. Across all eight regression models, the null hypothesis that $\beta$ equals zero cannot be rejected, indicating that the opponent's current choice does not significantly affect the player's decision in the current penalty scenario.

Interestingly, the results suggest that male kickers are somewhat influenced by their past actions, which may indicate a difficulty in achieving completely randomized actions. On the other hand, goalkeepers' decisions show a tendency to be influenced by the historical preferences of the kickers, as evidenced by significant coefficients in the analysis.

These findings are in line with those of Dohmen and Sonnabend (2018). They observe that while these results contradict the assumption of independence in choices over time, they do not necessarily negate the possibility of equilibrium play.

## 1.4 Testing the Mixed Strategy Nash Equilibrium

Once we have verified that players have similar preferences associated with the foot-type, and established that the assumptions of simultaneous-moves game and identical goalkeepers hold, we can proceed to check if players play the game following mixed strategy predictions.

The literature on testing mixed-strategy equilibria in soccer penalty kicks has predominantly utilized indirect methods (Chiappori et al.; 2002; Palacios-Huerta; 2003; Dohmen and Sonnabend; 2018). Their methodologies principally concentrate on determining whether players demonstrate equal scoring probabilities across their range of actions.

In contrast to these indirect methods, Coloma (2007) proposes a direct approach to test the mixed-strategy equilibrium, focusing on the specific results articulated by equations 1.4 to 1.9 for the 3x3 game (GR), and by equations 1.10 to 1.13 in case of the 2x2 game in which the center (C) is not an available option. This involves running a system of linear probability equations that relies on a fundamental homogeneity assumption of goalkeepers. The model aims to control for heterogeneity of the kickers by including variables related to the kicker's characteristics and match conditions, thereby isolating the average players' strategies and scoring probabilities.

In the case of the General Randomization (GR) game, the system is the following:

$$KNS = c_1 + \mathbf{X}\beta \tag{1.16}$$
$$KC = c_2 + \mathbf{X}\beta \tag{1.17}$$
$$KOS = c_3 + \mathbf{X}\beta \tag{1.18}$$
$$GNS = c_4 + \mathbf{X}\beta \tag{1.19}$$
$$GC = c_5 + \mathbf{X}\beta \tag{1.20}$$
$$GOS = c_6 + \mathbf{X}\beta \tag{1.21}$$

$$\begin{aligned} GOAL &= \gamma_1 KNS \cdot GNS + \gamma_2 KOS \cdot GOS \\ &+ \gamma_3 KC \cdot (1 - GC) + \gamma_4 KNS \cdot (1 - GNS) \\ &+ \gamma_5 KOS \cdot (1 - GOS) + \mathbf{X}\beta \end{aligned} \tag{1.22}$$

Conversely, for a RR equilibrium, the system comprises:

$$KNS = c_7 + \mathbf{X}\beta \tag{1.23}$$
$$KOS = c_8 + \mathbf{X}\beta \tag{1.24}$$
$$GNS = c_9 + \mathbf{X}\beta \tag{1.25}$$
$$GOS = c_{10} + \mathbf{X}\beta \tag{1.26}$$

$$\begin{aligned} GOAL &= \gamma_6 KNS \cdot GNS + \gamma_7 KOS \cdot GOS \\ &+ \gamma_8 KNS \cdot GOS + \gamma_9 KOS \cdot GNS + \mathbf{X}\beta \end{aligned} \tag{1.27}$$

Here, *KNS*, *KOS*, *KC*, *GNS*, *GOS*, and *GC* are dummy variables representing the actions of kickers (K) and goalkeepers (G). *GOAL* indicates whether the penalty kick resulted in a goal. The term **X** encompasses control variables for observable differences such as the kicker's foot type, home advantage, kicker's role, and whether the kicker's team is winning the game before the penalty kick. Equations 1.16 to 1.21 and 1.23 to 1.26 attempt to describe the strategies taken by the players controlling for the explanatory variables. The constants $c_1$ to $c_6$, represent the mixed strategies selected by players in the GR equilibrium. So that $c_1$, $c_2$ and $c_3$ are an estimation for $p_{NS}$, $p_C$ and $p_{OS}$, respectively; while $c_4$, $c_5$ and $c_6$ are the estimated probabilities $q_{NS}$, $q_C$ and $q_C$. Similarly, in a RR equilibrium, $c_7$, $c_8$, $c_9$ and $c_{10}$ are an estimation for $p_{NS}$, $p_{OS}$, $q_{NS}$, and $q_{OS}$, respectively. Equations 1.22 and 1.27 are designed to estimate the payoffs (or scoring probabilities) for the GR and RR games, respectively (see Tabs. 1.1 and 1.2).

In the GR equilibrium, when both players randomize over a set of three actions, $\gamma_1$ is an estimation for $\theta_{NS}$, $\gamma_2$ is an estimation for $\theta_{OS}$, $\gamma_3$ is an estimation for $\mu$, $\gamma_4$ is an estimation for $\pi_{NS}$, and $\gamma_5$ is an estimation for $\pi_{OS}$. Similarly, in a RR equilibrium, $\gamma_6$ is an estimation for $\theta_{NS}$, $\gamma_7$ is an estimation for $\theta_{OS}$, $\gamma_8$ is an estimation for $\pi_{NS}$, and $\gamma_9$ is an estimation for $\pi_{OS}$. Finally, after the estimation of the parameters, a joint *Wald test* is used to test the restrictions expressed by equations from 1.4 to 1.9 for the GR equilibrium and 1.10 to 1.13 for the RR equilibrium. For simplicity here are the constraints to be tested for the RR model:

$$c_7 = \frac{\gamma_9 - \gamma_7}{(\gamma_8 + \gamma_9 - \gamma_6 - \gamma_7)}, \tag{1.28}$$

$$c_8 = \frac{\gamma_8 - \gamma_6}{(\gamma_8 + \gamma_9 - \gamma_6 - \gamma_7)}, \tag{1.29}$$

$$c_9 = \frac{\gamma_8 - \gamma_7}{(\gamma_8 + \gamma_9 - \gamma_6 - \gamma_7)}, \tag{1.30}$$

and

$$c_{10} = \frac{\gamma_9 - \gamma_6}{(\gamma_8 + \gamma_9 - \gamma_6 - \gamma_7)}, \tag{1.31}$$

whereas the constraints of the 3x3 model are given in the Appendix [Equations 1.32 to 1.37]

### 1.4.1 Estimation and Results

Table 1.7 reports the estimated frequencies (constants of the equations) alongside the payoff matrix parameters of the GR model (see Table 1.1). The regression results indicate that most of the parameters are highly statistically significant. A notable difference between male and female results exists in the statistical significance of coefficients related to the Center actions. Indeed, the estimated $q_C$ for female goalies is not significantly different from zero, suggesting a divergence from expected strategic behavior in these scenarios.

Assuming that players of both genders participate in the 3x3 game, the implications of the GR Proposition [see Eq. 1.1] suggest a high value of $\mu$. When we apply the Wald test to the GR Proposition, notable gender differences emerge. For female players, $\mu$ does not significantly exceed the right-hand side of the proposition, as indicated by a $\chi^2$ value of 1.4751 and a p-value = 0.1123. In contrast, the test of the GR Preposition for male players suggests a greater alignment with the 3x3 game model ($\chi^2$ = 3.078, p-value = 0.036).

Table 1.7 also includes the $R^2$ values for each regression. A limitation of our analysis is the low $R^2$ values for the KNS, KOS, GNS, and GOS equations. This issue arises from using linear probability models with dummy variables as dependent variables. The primary objective here is not to predict values for these variables' different observations but to estimate more precise probabilities $p_S$ and $q_S$, with $S = \{N, C, O\}$, and the payoffs $\theta_{NS}$, $\theta_{OS}$, $\mu$,$\pi_{NS}$, $\pi_{OS}$.

By converting our regression estimates into a normal-form game format (as shown in Tables 1.8 and 1.9), distinct strategic patterns emerge between genders. While male players' actual frequencies seem to follow almost perfectly the MSNE predictions, female ones are somehow different, especially those relating to goalkeepers. Specifically, given their respective payoffs, female goalkeepers should theoretically play choice C with a negative probability, which is, of course, an impossibility. This outcome is a result of the rejection of the GR proposition [Eq. 1.1].

**Table 1.7:** *Regression Results - GR model*

| | Females | | | Males | | |
|---|---|---|---|---|---|---|
| *Equation* | Coeff. | SE | p | Coeff | SE | p |
| *KNS Equation* | | | | | | |
| Constant ($p_{NS}$) | 0.540 | 0.046 | 0.000 | 0.488 | 0.027 | 0.000 |
| $R^2$ | 0.002 | | | 0.004 | | |
| *KC Equation* | | | | | | |
| Constant ($p_C$) | 0.054 | 0.024 | 0.025 | 0.152 | 0.020 | 0.000 |
| $R^2$ | 0.004 | | | 0.002 | | |
| *KOS Equation* | | | | | | |
| Constant ($p_{OS}$) | 0.406 | 0.045 | 0.000 | 0.360 | 0.026 | 0.000 |
| $R^2$ | 0.004 | | | 0.002 | | |
| *GNS Equation* | | | | | | |
| Constant ($q_{NS}$) | 0.556 | 0.046 | 0.000 | 0.547 | 0.027 | 0.000 |
| $R^2$ | 0.007 | | | 0.001 | | |
| *GC Equation* | | | | | | |
| Constant ($q_C$) | 0.024 | 0.020 | 0.238 | 0.037 | 0.011 | 0.001 |
| $R^2$ | 0.006 | | | 0.001 | | |
| *GOS Equation* | | | | | | |
| Constant ($q_{OS}$) | 0.420 | 0.045 | 0.000 | 0.416 | 0.026 | 0.000 |
| $R^2$ | 0.008 | | | 0.001 | | |
| *GOAL Equation* | | | | | | |
| Parameter $\theta_{NS}$ | 0.566 | 0.041 | 0.000 | 0.626 | 0.022 | 0.000 |
| Parameter $\theta_{OS}$ | 0.559 | 0.048 | 0.000 | 0.549 | 0.026 | 0.000 |
| Parameter $\mu$ | 0.655 | 0.064 | 0.000 | 0.800 | 0.025 | 0.000 |
| Parameter $\pi_{NS}$ | 0.871 | 0.044 | 0.000 | 0.936 | 0.024 | 0.000 |
| Parameter $\pi_{OS}$ | 0.893 | 0.044 | 0.000 | 0.955 | 0.024 | 0.000 |
| $R^2$ | 0.173 | | | 0.174 | | |
| N | 911 | | | 3386 | | |

The joint Wald tests on the MSNE-related constraints show that for men, the null hypothesis of strategy probabilities being equal to those predicted by MSNE, given the payoffs, is not rejected ($\chi^2(6) = 2.072; p = 0.913$). Conversely, for women, the null hypothesis of using mixed strategies in a 3x3 game is rejected at a 5% significance level. This highlights a significant gender difference in the adoption of mixed strategies in the 3x3 Penalty Kick game.

**Table 1.8:** *3x3 game: Female players*

| | Goalkeeper | | | Actual | MSNE |
|---|---|---|---|---|---|
| Kicker | NS | C | OS | freq. | pred. |
| NS | 0.566 | 0.871 | 0.871 | 0.540 | 0.47 |
| C | 0.655 | 0 | 0.655 | 0.054 | 0.08 |
| OS | 0.893 | 0.893 | 0.559 | 0.406 | 0.43 |
| Actual freq. | 0.556 | 0.024 | 0.420 | | |
| MSNE pred. | 0.53 | -0.08 | 0.55 | | |
| Wald test | df = 6 | $\chi^2 = \mathbf{16.186}$ | | p-value = **0.0127** | |

Given these findings, it becomes evident that, in contrast to men, women players seemingly discount the center as a viable strategic option when playing the game. This observation leads us to continue the analysis, exploring the dynamics of the 2x2 game framework in which the center (C) is categorized as a component of the Natural Side strategy (NS) for both players, as hypnotized by Palacios-Huerta (2003). The results of the restricted model are shown in Table 1.10.

The linear probability model system yields significant estimates across all equations about both the probabilities and the payoffs for both the kicker and the goalie. This consistency in

**Table 1.9:** *3x3 game: Male players*

| | Goalkeeper | | | Actual | MSNE |
|---|---|---|---|---|---|
| Kicker | NS | C | OS | freq. | pred. |
| NS | 0.626 | 0.936 | 0.936 | 0.516 | 0.488 |
| C | 0.800 | 0 | 0.800 | 0.089 | 0.152 |
| OS | 0.955 | 0.955 | 0.549 | 0.394 | 0.360 |
| Actual freq. | 0.556 | 0.024 | 0.420 | | |
| MSNE pred. | 0.547 | 0.037 | 0.416 | | |
| Wald test | df = 6 | $\chi^2 = \mathbf{2.0721}$ | | p-value = **0.9129** | |

**Table 1.10:** *Regression Results - RR model where C = NS*

| | Females | | | Males | | |
|---|---|---|---|---|---|---|
| *Equation* | Coeff. | SE | p | Coeff | SE | p |
| *KNS Equation* | | | | | | |
| Constant | 0.594 | 0.045 | 0.000 | 0.640 | 0.026 | 0.000 |
| $R^2$ | 0.004 | | | 0.002 | | |
| *KOS Equation* | | | | | | |
| Constant | 0.406 | 0.045 | 0.000 | 0.360 | 0.026 | 0.000 |
| $R^2$ | 0.004 | | | 0.002 | | |
| *GNS Equation* | | | | | | |
| Constant | 0.580 | 0.045 | 0.000 | 0.584 | 0.026 | 0.000 |
| $R^2$ | 0.008 | | | 0.001 | | |
| *GOS Equation* | | | | | | |
| Constant | 0.420 | 0.045 | 0.000 | 0.416 | 0.026 | 0.000 |
| $R^2$ | 0.008 | | | 0.001 | | |
| *GOAL Equation* | | | | | | |
| Parameter $\theta_{NS}$ | 0.577 | 0.041 | 0.000 | 0.648 | 0.022 | 0.000 |
| Parameter $\theta_{OS}$ | 0.559 | 0.049 | 0.000 | 0.539 | 0.027 | 0.000 |
| Parameter $\pi_{NS}$ | 0.835 | 0.045 | 0.000 | 0.933 | 0.025 | 0.000 |
| Parameter $\pi_{OS}$ | 0.892 | 0.046 | 0.000 | 0.949 | 0.024 | 0.000 |
| $R^2$ | 0.133 | | | 0.147 | | |
| N | 911 | | | 3386 | | |

*Notes:* The notation KNS (GNS) means that the dependent variable was a dummy for whether the kicker (goalkeeper) chooses the Natural Side (NS) or Center (C)

significance across genders and strategic choices underscores a more uniform application of game strategies within the 2x2 game framework compared to the 3x3 game context.

Tables 1.11 and 1.12 present the outcomes of the system's estimates structured in the typical normal form of the game. The outcomes of the joint Wald tests on the mixed strategies constraints are markedly supportive of the hypothesis that the observed data reflects the Mixed Strategy Nash Equilibrium prediction.

**Table 1.11:** *2x2 game: Female players*

| | Goalkeeper | | Actual | MSNE |
|---|---|---|---|---|
| Kicker | NS | OS | freq. | predict |
| NS | 0.577 | 0.835 | 0.594 | 0.564 |
| OS | 0.892 | 0.559 | 0.406 | 0.435 |
| Actual freq. | 0.580 | 0.420 | | |
| MSNE pred. | 0.467 | 0.532 | | |
| Wald test | df = 4 | $\chi^2 = \mathbf{3.8368}$ | p-value = **0.4285** | |

**Table 1.12:** *2x2 game: Male players*

| Kicker | Goalkeeper NS | Goalkeeper OS | Actual freq. | MSNE predict |
|---|---|---|---|---|
| NS | 0.648 | 0.933 | 0.640 | 0.590 |
| OS | 0.949 | 0.539 | 0.360 | 0.410 |
| Actual freq. | 0.584 | 0.416 | | |
| MSNE pred. | 0.567 | 0.433 | | |
| Wald test | df = 4 | $\chi^2$ = **2.3915** | p-value = **0.6642** | |

### 1.4.2 Re-evaluating the Center option in Penalty Kicks

Our analysis reveals a distinctive trend among female players to avoid choosing the center option in penalty kicks. This finding challenges the assumption made by Palacio-Huerta, which suggests that players, particularly professionals, consider kicking toward the center the same as their natural side strategy. Palacio-Huerta's conclusion was partly based on personal interviews with professional players, who indicated a preference for using the interior side of the foot for better control, thereby treating C and NS options as effectively equivalent.

The divergence observed in our study between the actual in-game behavior of female players in using the center option and the theoretical assumption posited by Palacios-Huerta (2003) raises important questions. It suggests that the reality of gameplay, especially in high-stakes situations like penalty kicks, might not always align with theoretical models. Given these insights, we conducted a further analysis by redefining the strategic framework. In this revised approach we treat the center option as the Opposite Side (OS). Results of the system of linear equations are provided in Table 1.13. Like before, the regressions bring statistically significant estimates for both genders.

Again, our analysis extends to examining the game's structure in its normal form, as presented in Tables 1.14 and 1.15. Notably, for female players, the dynamics of the game closely resemble those of the Matching Pennies game, in which players to randomize have to assign equal probabilities to each strategy[7]. Indeed, according to the MSNE predictions, both players – the kicker and the goalkeeper – should ideally assign a probability close to 50% to each option to align precisely with theoretical predictions. The joint Wald test results (p=0.8636) further reinforce the idea that the data could have been generated by mixed strategies play, with the center seen as the non-natural side OS. The same could be said for men, for whom the joint Wald test on the RR model with C = OS improves from the previous one, but remains lower than the probability that the randomization occurs on a 3x3 game.

Our findings imply that while the use of the interior side of the foot might offer more control, this does not necessarily translate into a uniform strategy of equating the center with the natural side. This could be particularly true for female players who, as our data suggest, exhibit a distinct strategic pattern that avoids randomization over C.

[7] The Matching Pennies game is a fundamental example in game theory, often used to illustrate the concept of mixed strategies and zero-sum games. The only equilibrium in this game is a mixed strategy equilibrium where each player randomly chooses Heads or Tails with equal probability (50%).

**Table 1.13:** *Regression Results - RR model where C = OS*

| | Females | | | Males | | |
|---|---|---|---|---|---|---|
| *Equation* | Coeff. | SE | p | Coeff | SE | p |
| *KNS Equation* | | | | | | |
| Constant | 0.540 | 0.046 | 0.000 | 0.488 | 0.027 | 0.000 |
| $R^2$ | 0.002 | | | 0.004 | | |
| *KOS Equation* | | | | | | |
| Constant | 0.460 | 0.046 | 0.000 | 0.512 | 0.027 | 0.000 |
| $R^2$ | 0.002 | | | 0.004 | | |
| *GNS Equation* | | | | | | |
| Constant | 0.556 | 0.046 | 0.000 | 0.547 | 0.027 | 0.000 |
| $R^2$ | 0.009 | | | 0.001 | | |
| *GOS Equation* | | | | | | |
| Constant | 0.444 | 0.046 | 0.000 | 0.453 | 0.027 | 0.000 |
| $R^2$ | 0.009 | | | 0.001 | | |
| *GOAL Equation* | | | | | | |
| Parameter $\theta_{NS}$ | 0.566 | 0.042 | 0.000 | 0.627 | 0.024 | 0.000 |
| Parameter $\theta_{OS}$ | 0.562 | 0.047 | 0.000 | 0.642 | 0.025 | 0.000 |
| Parameter $\pi_{NS}$ | 0.871 | 0.046 | 0.000 | 0.937 | 0.025 | 0.000 |
| Parameter $\pi_{OS}$ | 0.866 | 0.045 | 0.000 | 0.904 | 0.024 | 0.000 |
| $R^2$ | 0.140 | | | 0.124 | | |
| N | 911 | | | 3386 | | |

*Notes:* The notation KOS (GOS) means that the dependent variable was a dummy for whether the kicker (goalkeeper) chooses the Opposite Side (OS) or Center (C)

**Table 1.14:** *2x2 game: Female players*

| | Goalkeeper | | Actual | MSNE |
|---|---|---|---|---|
| Kicker | NS | OS | freq. | predict |
| NS | 0.566 | 0.871 | 0.564 | 0.499 |
| OS | 0.866 | 0.562 | 0.460 | 0.501 |
| Actual freq. | 0.556 | 0.444 | | |
| MSNE pred. | 0.506 | 0.493 | | |
| Wald test | df = 4 | $\chi^2$ = **1.2867** | p-value = **0.8636** | |

**Table 1.15:** *2x2 game: Male players*

| | Goalkeeper | | Actual | MSNE |
|---|---|---|---|---|
| Kicker | NS | OS | freq. | predict |
| NS | 0.627 | 0.937 | 0.488 | 0.458 |
| OS | 0.904 | 0.642 | 0.512 | 0.541 |
| Actual freq. | 0.547 | 0.453 | | |
| MSNE pred. | 0.515 | 0.485 | | |
| Wald test | df = 4 | $\chi^2$ = **1.9459** | p-value = **0.7457** | |

## 1.5 Discussion and Concluding Remarks

This study aimed to explore mixed strategies in natural experiments, with a specific focus on gender differences in high-pressure situations exemplified by the Penalty Kick game. Penalty kicks provide a unique opportunity to analyze decision-making and strategic choices in a stress-induced environment since often they can change the equilibrium of a match or even the victory of the competition.

Utilizing an extensive dataset on penalty kicks, our investigation affirmed the theoretical predictions regarding playing behaviors under such pressure. Descriptive statistics demonstrated similarities in the probability of scoring under various conditions - including score advantages, roles of the kicker, and foot type - suggesting that both genders exhibit comparable behaviors in these high-stress moments of their careers as players.

Then we successfully tested that the Penalty Kick game is a simultaneous moves game where players of both genders have the same preferences linked with their preferred foot to kick the ball.

Since only rarely did the same kicker face the same goalkeeper, an analysis through a real experiment between the same players was not possible. Instead, we tested the hypothesis of identical goalkeepers previously used in literature to check if goalkeepers can be treated as homogeneous during the penalty kicks. In none of the cases, we could reject the hypotheses

After having outlined the methodological model that was used to estimate the probabilities of choices and payoffs, it was tested on three different scenarios. The first assumed that both genders were randomized using mixed strategies on a 3x3 game, the GR model. Afterward, 2x2 games were tested in which the central choice was first taken as part of the Natural Side and then the Opposite Side.

The results of the joint Wald tests on the theoretical restrictions of the MSNE have brought about very interesting results. Female players showed notable avoidance of the middle option C, leading to rejection of the hypothesis that the data is consistent with a 3x3 game model. Instead, this behavior aligns more with a 2x2 gameplay model. In this last case, unlike what has been established so far by the literature on penalty kicks, both the samples of women and men seem more in line with the hypothesis that the central option can be seen as a non-natural side. Nonetheless, men seem to follow the theory remarkably in the 3x3 game, as already seen in the literature Chiappori et al. (2002); Palacios-Huerta (2003); Azar and Bar-Eli (2011).

Notably, despite these strategic differences, the final performance outcomes between male and female players do not show a statistically significant divergence. This observation highlights that while strategic preferences may vary, they do not necessarily impact the effectiveness of performance in penalty kicks across genders.

The divergent strategic choices observed in penalty kicks across genders can be partially explained by the *Action bias* detailed by Bar-Eli et al. (2007). They identified an Action bias - the opposite of the more famous 'Omission bias' - in goalkeepers, who are more inclined to jump rather than stay in the center during penalty kicks. This inclination is tied to *Norm Theory* elaborated by Kahneman and Miller (1986), which suggests that individuals perceive higher disutilities from negative outcomes resulting from abnormal causes or decisions. In penalty kicks, action - jumping or kicking to one of the sides - is considered the norm. This bias towards action may indicate a psychological preference for making a visible effort, driven by the belief that active efforts to stop the penalty will face less criticism and regret, even if unsuccessful. It highlights the thought processes of players in high-pressure situations, where taking action is favored over inaction.

Incorporating this concept into our analysis, it appears that female players may be influenced by an action bias. The intense pressure of the situation, potentially heightened by the crowd's presence and the high stakes, might lead them to prefer actions that are perceived as more 'active' or 'decisive'. This tendency to choose more dynamic actions sheds light on the complex thought processes players go through when under pressure.

The concept of 'action bias' could be extended beyond the penalty kick scenario, revealing broader economic and managerial implications that mirror the differences in decision-making strategies observed between female and male players. The inclination towards action over inaction, as highlighted by the norm theory, suggests a profound psychological underpinning that influences not just athletes but also investors, managers, workers, and policymakers. In economic and managerial decisions, for example, the tendency to prefer action over omission could influence the behavior of investors in portfolio management, the strategic decisions of company leaders, the job-seeking behavior of employees, and even the policy-making processes of governments and central banks. Each of these scenarios reflects a context where the perceived need for decisive action might override the potential benefits of inaction or a more measured approach.

Despite these intriguing insights, it is not possible to attribute with certainty that these gender differences in playing the Penalty Kick game are generated by different degrees of risk perception due to an action bias. A limitation of this study is, therefore, the absence of a model capable of accurately estimating the intensity of risk aversion of players involved.

Nevertheless, professional sports are interesting contexts for studying game theory solution concepts such as the Mixed-strategy Nash Equilibrium. The competitiveness of the players, their preparation for the game, and the huge stakes contribute to making professional sports players very similar to those theorized by economic theory.

## References


Azar, O. H. and Bar-Eli, M. (2011). Do soccer players play the mixed-strategy Nash equilibrium?, *Applied Economics* **43**(25): 3591–3601.

Bar-Eli, M., Azar, O. H., Ritov, I., Keidar-Levin, Y. and Schein, G. (2007). Action bias among elite soccer goalkeepers: The case of penalty kicks, *Journal of Economic Psychology* **28**(5): 606–621.

Brown, J. N. and Rosenthal, R. (1990). Testing the minimax hypothesis: A re-examination of o'neill's game experiment, *Econometrica* **58**(5): 1065–81.

Charness, G. and Gneezy, U. (2012). Strong Evidence for Gender Differences in Risk Taking, *Journal of Economic Behavior and Organization* **83**(1): 50–58.

Chiappori, P. A., Levitt, S. and Groseclose, T. (2002). Testing mixed-strategy equilibria when players are heterogeneous: The case of penalty kicks in soccer, *American Economic Review* **92**(4): 1138–1151.

Coloma, G. (2007). Penalty Kicks in Soccer: An Alternative Methodology for Testing Mixed-Strategy Equilibria, *Journal of Sports Economics* **8**(5): 530–545.

Dixit, A. K., Skeath, S. and McAdams, D. (2020). *Games of Strategy: Fifth International Student Edition*, WW Norton & Company.

Dohmen, T. and Sonnabend, H. (2018). Further field evidence for minimax play, **19**(3): 371–388.

Eckel, C. C. and Grossman, P. J. (2008). Men, women and risk aversion: Experimental evidence, *Handbook of experimental economics results* **1**: 1061–1073.

Gauriot, R., Page, L. and Wooders, J. (2023). Expertise, gender, and equilibrium play, *Quantitative Economics* **14**(3): 981–1020.

Gneezy, U., Niederle, M. and Rustichini, A. (2003). Performance in competitive environments: Gender differences, *Quarterly Journal of Economics* **118**(3): 1049–1074.

Goeree, J. K., Holt, C. A. and Palfrey, T. R. (2003). Risk averse behavior in generalized matching pennies games, *Games and Economic Behavior* **45**(1): 97–113.

Jeon, G. R. and Ong, D. (2020). The Gender Difference in Mixed Strategy Nash Equilibrium Play, (855).

Kahneman, D. and Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives., *Psychological Review* **93**(2): 136–153.

O'Neill, B. (1987). Nonmetric test of the minimax theory of two-person zerosum games., *Proceedings of the National Academy of Sciences of the United States of America* **84**(7): 2106–2109.

Palacios-Huerta, I. (2003). Professionals play minimax, *Review of Economic Studies* **70**(2): 395–415.

Paserman, M. D. (2011). Gender Differences in Performance in Competitive Environments: Evidence from Professional Tennis Players, *SSRN Electronic Journal* (2834).

Rapoport, A. and Budescu, D. V. (1992). Generation of random series in two-person strictly competitive games., *Journal of Experimental Psychology: General* **121**(3): 352–363.

Shachat, J. M. (2002). Mixed strategy play and the minimax hypothesis, *Journal of Economic Theory* **104**(1): 189–226.

Walker, M. and Wooders, J. (2001). Minimax play at wimbledon, *American Economic Review* **91**(5): 1521–1538.

# Appendix

The restrictions tested by the Wald test in the GR model are:

$$p_{NS} = \frac{\gamma_3 \cdot (\gamma_5 - \gamma_2)}{(\gamma_5 - \gamma_2) \cdot (\gamma_4 - \gamma_1) + \gamma_3 \cdot (\gamma_4 + \gamma_5 - \gamma_1 - \gamma_2)}, \tag{1.32}$$

**Table A1:** *Foot-types and Preferences of Goalkeepers*

| | Foot-type | Jump side: Left | Center | Right | Total |
|---|---|---|---|---|---|
| Females | Left-footed | 57 *40.4%* | 6 *4.3%* | 78 *55.3%* | 141 *100 %* |
| Females | Right-footed | 461 *59.9%* | 38 *4.9 %* | 271 *35.2%* | 770 *100 %* |
| Females | Total | 518 *56.9%* | 44 *4.8 %* | 349 *38.3%* | 911 *100 %* |
| Males | Left-footed | 267 *36.1%* | 29 *3.9%* | 443 *59.9%* | 739 *100 %* |
| Males | Right-footed | 1539 *58.1%* | 111 *4.2 %* | 997 *37.5%* | 2647 *100 %* |
| Males | Total | 1806 *53.3%* | 140 *4.1 %* | 1440 *42.4%* | 3386 *100 %* |

*Notes*: $\chi^2_{\text{female}}(2) = 20.637, p = 0.000$; $\chi^2_{\text{male}}(2) = 120.015, p = 0.000$.

**Table A2:** *Joint significance of goalie-fixed effects across different restrictions*

| | | | | Df | | **Saved** | | **KNS** | | **GNS** | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Restr. | N. Obs. | N. GK. | Num. | Den. | F value | *Pr(>F)* | F value | *Pr(>F)* | F value | *Pr(>F)* |
| Females | All | 911 | 207 | 185 | 341 | 1.291 | 0.022* | 0.849 | 0.894 | 1.035 | 0.389 |
| | 2+ | 855 | 151 | 143 | 341 | 1.304 | 0.027* | 0.795 | 0.943 | 1.079 | 0.287 |
| | 3+ | 775 | 111 | 106 | 322 | 1.227 | 0.091 | 0.822 | 0.883 | 1.116 | 0.234 |
| | 4+ | 694 | 84 | 79 | 293 | 1.168 | 0.181 | 0.750 | 0.936 | 1.058 | 0.363 |
| | 5+ | 642 | 71 | 67 | 272 | 1.190 | 0.170 | 0.744 | 0.926 | 1.139 | 0.235 |
| | 6+ | 582 | 59 | 56 | 240 | 1.186 | 0.193 | 0.659 | 0.968 | 1.264 | 0.118 |
| | 7+ | 522 | 49 | 48 | 209 | 1.111 | 0.303 | 0.754 | 0.878 | 1.270 | 0.130 |
| | 8+ | 438 | 37 | 36 | 153 | 1.207 | 0.216 | 1.018 | 0.451 | 1.412 | 0.079 |
| | 9+ | 342 | 25 | 23 | 111 | 1.192 | 0.268 | 0.459 | 0.983 | 1.218 | 0.245 |
| | 10+ | 324 | 23 | 22 | 105 | 1.161 | 0.299 | 0.465 | 0.979 | 1.375 | 0.144 |
| Males | All | 3386 | 395 | 375 | 2082 | 0.967 | 0.658 | 0.919 | 0.851 | 1.005 | 0.467 |
| | 2+ | 3309 | 318 | 314 | 2082 | 1.015 | 0.423 | 0.874 | 0.937 | 0.962 | 0.664 |
| | 3+ | 3207 | 267 | 266 | 2044 | 0.971 | 0.616 | 0.906 | 0.850 | 0.959 | 0.667 |
| | 4+ | 3105 | 233 | 232 | 1992 | 1.021 | 0.406 | 0.924 | 0.779 | 1.001 | 0.488 |
| | 5+ | 2985 | 203 | 202 | 1924 | 1.035 | 0.359 | 0.876 | 0.887 | 0.996 | 0.504 |
| | 6+ | 2885 | 183 | 182 | 1858 | 1.054 | 0.305 | 0.964 | 0.620 | 1.070 | 0.259 |
| | 7+ | 2807 | 170 | 168 | 1807 | 1.050 | 0.321 | 0.975 | 0.577 | 1.043 | 0.344 |
| | 8+ | 2688 | 153 | 151 | 1719 | 1.089 | 0.227 | 1.012 | 0.448 | 1.053 | 0.321 |
| | 9+ | 2624 | 145 | 143 | 1670 | 1.044 | 0.351 | 1.028 | 0.398 | 1.080 | 0.253 |
| | 10+ | 2471 | 128 | 126 | 1564 | 1.094 | 0.232 | 1.037 | 0.377 | 1.067 | 0.296 |

*Notes*: The restrictions refer to the number of penalty kicks for each goalkeeper available in our sample. If a goalkeeper is homogeneous, the *F test* should not reject the *null hypothesis* that all goalie-fixed effects are equal. All regressions also include the control variables for whether the kicker was kicking the penalty *home*, which was the *score difference* before the observed penalty, the *quarter* in which the penalty was taken, and the *role* of the kicker. $^{*}$p<0.05; $^{**}$p<0.01; $^{***}$p<0.001

$$p_C = \frac{(\gamma_5 - \gamma_2) \cdot (\gamma_4 - \gamma_1)}{(\gamma_5 - \gamma_2) \cdot (\gamma_4 - \gamma_1) + \gamma_3 \cdot (\gamma_4 + \gamma_5 - \gamma_1 - \gamma_2)}, \tag{1.33}$$

$$p_{OS} = \frac{\gamma_3 \cdot (\gamma_4 - \gamma_1)}{(\gamma_5 - \gamma_2) \cdot (\gamma_4 - \gamma_1) + \gamma_3 \cdot (\gamma_4 + \gamma_5 - \gamma_1 - \gamma_2)}, \tag{1.34}$$

$$q_{NS} = \frac{\gamma_4 \cdot (\gamma_5 - \gamma_2) + \gamma_3 \cdot (\gamma_4 - \gamma_1)}{(\gamma_5 - \gamma_2) \cdot (\gamma_4 - \gamma_1) + \gamma_3 \cdot (\gamma_4 + \gamma_5 - \gamma_1 - \gamma_2)}, \tag{1.35}$$

$$q_C = \frac{\gamma_3 \cdot (\gamma_4 + \gamma_5 - \gamma_1 - \gamma_2) + \gamma_1 \cdot \gamma_2 - \gamma_4 \cdot \gamma_5}{(\gamma_5 - \gamma_2) \cdot (\gamma_4 - \gamma_1) + \gamma_3 \cdot (\gamma_4 + \gamma_5 - \gamma_1 - \gamma_2)}, \tag{1.36}$$

and

$$q_{OS} = \frac{\gamma_5 \cdot (\gamma_4 - \gamma_1) - \gamma_3 \cdot (\gamma_4 - \gamma_1)}{(\gamma_5 - \gamma_2) \cdot (\gamma_4 - \gamma_1) + \gamma_3 \cdot (\gamma_4 + \gamma_5 - \gamma_1 - \gamma_2)} \tag{1.37}$$

# Chapter 2

# The Pursuit of Fairness in Ongoing Evaluations: Evidence from Referees in Soccer


## Abstract

This study investigates the presence of a fairness bias in referees' decision-making across various scenarios in professional soccer matches. By analyzing a comprehensive dataset from the top five European leagues over 10 seasons, we uncover systematic evidence of referees' tendencies to balance their rulings, aiming to reduce perceived inequity between competing teams. This research is grounded in behavioral economics theories, with a particular focus on inequity aversion—a concept where individuals demonstrate a preference for fair outcomes and may alter their behavior to achieve perceived equity. Our findings reveal significant instances of even-out bias in almost all scenarios requiring a referee's decision, such as awarding penalty kicks, issuing red and yellow cards, and allocating the injury time. These findings indicate that referees' decisions are influenced not only by the immediate play but also by a subconscious drive to rectify perceived past decisions. The implications of this study are profound, enhancing our understanding of referees' behavior and highlighting the impact of non-monetary incentives on the decision-making processes of individuals.



**JEL Classification:** C93, D91
**Keywords:** Behavioral Economics; Inequity Aversion; Decision-making; Even-out Bias; Referee's Behaviour.

## 2.1 Introduction

The economic literature has extensively demonstrated the crucial role of monetary incentives on individual behavior. However, it is equally intriguing to examine the effects of non-monetary incentives. In this context, sports events, owing to the practicality of data sources available, have become a fertile ground for numerous studies. A concrete example of a non-monetary incentive is the influence that the audience at a sporting event exerts on refereeing decisions.

For instance, Garicano et al. (2005) investigates the impact of social pressure on referees. This pressure can 'corrupt' referees decisions, leading to favoritism towards the home team. One observed behavior is the extension of game duration when the home team is losing. Scoppa (2021) utilized the distinctive scenario of empty stadiums during the COVID-19 pandemic in 2020 to conduct a natural experiment. His study revealed that the absence of spectators significantly impacts both teams and referees regarding performance and decision-making. Several other studies have confirmed the home advantage bias in soccer, (Boyko et al.; 2007; Dohmen; 2008; Dohmen and Sauermann; 2016). These findings contribute to the understanding that individual behavior is influenced not only by monetary incentives but also by social and psychological rewards.

Beyond external factors like audience presence, referees' judgment may also be influenced by internal non-monetary incentives. A key example is their aspiration to maintain a role of impartiality in the game. This pursuit of fairness in the game's outcome could lead referees to unconsciously seek to equalize the decisions made during the match. For example, Plessner and Betsch (2001), using videotaped scenes from actual matches, demonstrate that referees' decisions are impacted by preceding events in a match. Their study on penalty kick decisions reveals a tendency to avoid awarding successive penalties to the same team, indicating a compensatory bias or an inclination towards equalizing decisions within a game.

While controlled laboratory studies provide valuable insights, their applicability to real-world refereeing contexts is debated. Mascarenhas et al. (2002) argue that the natural decision-making environment of referees, characterized by high-stakes, real-time judgments often made under intense pressure, differs significantly from the more controlled, spectator-view settings of laboratory experiments. Building upon this framework, Schwarz (2011) further investigates compensatory biases in soccer referees' decisions. His analysis of matches from the German Bundesliga highlights a tendency among referees to balance penalty decisions across teams, suggesting an underlying effort to 'balance things out' and maintain fairness in the game. In line with this, Anderson and Pierce (2009) analyzed foul calls in NCAA basketball games and found a propensity in referees to balance foul calls between teams, with a notable bias against visiting teams and teams in the lead. These findings suggest a broader behavioral pattern among referees to maintain perceived fairness, potentially at the cost of objective decision-making. This need for referees to necessarily maintain a balanced behavior in decisions could be traced back to the model of *inequity aversion* discussed by Fehr and Schmidt (1999).

Inequity aversion implies that individuals often avoid outcomes perceived as unfair, even at their expense. This behavior is in contrast with most theoretical models that assume that individuals are rational actors moved only by material self-interests in their decision-making. This is only one hypothesis. Another explanation could be due to their aversion to influencing the final result due to a wrong call. In his book *Scorecasting*, Wertheim (2011) hypothesizes that "When an obviously bad call is made, the officials soon compensate 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"[1].

Further, another motivation suggested by Schwarz (2011) is 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 area or simply appealing to the referee's impartiality, inducing them to make an erroneous call that evens the score. More simply, the even-out bias of referees could be a fundamental trait deemed right to maintain peace among players on the field Noecker and Roback (2012).

Furthermore, as demonstrated by Haynes and Gilovich (2010), even players might show a sort of even-out bias. They find that basketball players tend to miss more free throws when these are the result of incorrect calls by the referee, indicating their conflicting behavior in inequity situations. The same researchers find comparable results in tennis; in this case, the perceived unfair advantage does not stem from a referee's call but simply from luck. That is, when a player wins a point during a tennis match thanks to a lucky bounce on the net, the same player wins the next point only 29% of the time, much lower than the 50% expected by chance.

The concept of even-out bias extends beyond the realm of sport. It mirrors a broader pattern of human behavior, where decision-makers balance choices for perceived fairness. Indeed, the difference between creating an impression of fairness and behaving fairly has potential effects in many fields, particularly in any kind of assessment or evaluation. Teachers may grade some students harder or easier than others for the sake of avoiding conflict and encouraging each student to learn, even if that means departing from a completely fair grading scale. Or again, what seems like fair compensation to an employee may not correspond with what is fair based strictly on achievement, a disparity that can potentially have real positive or negative economic effects.

Given the complexity and multidimensional nature of referees' decision-making, this study aims to delve deeper. We integrate existing literature findings with an in-depth investigation of referees' even-out bias. This research seeks to provide a comprehensive understanding of how non-monetary incentives and psychological factors influence referees' decisions, particularly in the context of professional soccer.

The paper is structured as follows. Section 2 will present the data and initial statistical analysis. Section 3 then delves into the methodology, elucidating the statistical methods and analytical approaches used to dissect referees' decision-making processes. Section 4 presents the evidence we find on even-out bias concerning penalty kicks, red and yellow cards. Lastly, sections 5 and 6 provide a discussion of the results and the implications for policy-making and future research directions.

[1] An example is reported by the *Corriere dello Sport*. When an Italian soccer referee was asked why he did not call a clear penalty during a Serie A match, the response was "Because I didn't call the one for the other team."

## 2.2 Data and Descriptive Statistics

The data encompasses 18007 matches from the top five European soccer leagues over ten seasons, from 2012-13 to 2022-23. Table B1 in the Appendix outlines the dataset distribution, categorizing the matches by season and league. The choice of these specific leagues and seasons ensures a high level of competitiveness and allows for a robust analysis of referee behavior across diverse settings.

The methodology of this study centers on an in-depth analysis of soccer-match commentary generated from live play, which is sourced from *ESPN.com.* From these commentaries we meticulously extracted all necessary information relevant to the match, together with the minute in which they occurred. This process allowed us to analyze specific events and actions within their exact temporal context [2]. Soccer matches often extend beyond 90 minutes due to injury time. Our methodological approach, therefore, partitions each match into $T = 100$ equal time frames. This follows the percentile-based segmentation proposed by Robberechts et al. (2019). This temporal division ensures that each segment of the game is equally represented, with the end of the first half consistently aligned with the 50th percentile, facilitating the analysis of time-dependent variables. This approach is important for capturing the dynamic nature of soccer matches and referee decisions during a game. Indeed, team managers frequently adapt the strategies of their teams, adjusting both their attacking and defensive tactics. These adjustments include changes in players, with substitutions that can vary in quality, affecting the team's performance.

Additional to this time-frame specif observations, we collected additional data regarding match statistics, audience attendance, and referee identity from *FBstat.com* and *transfermarkt.com*, together with weekly values of the Elo scores from *clubelo.com*[3].

Table 2.1 summarizes the descriptive statistics dividing the variables in *Time frame variables* and *Match variables*.

Given the division into 100 time frames per game, our dataset results in a total of 1800700 observations. Given the structure of the data, mean values for the time frame variables are relatively low; however, it is still observable that home teams appear to be consistently advantaged. Indeed, the home teams score more goals ($t = 27.246$), commit fewer fouls ($t = -7.0229$), reach more shooting opportunities ($t = 64.376$), and receive fewer disciplinary sanctions: fewer yellow cards ($t = -16.513$), and red cards ($t = -7.9698$). It is important to take into account that around 40% of the red cards are the result of a player receiving a second yellow card during the same match, which leads to the automatic red card. Lastly, they are awarded more penalty kicks ($t = 12.328$). Conversely, there is no significant difference in the number of injuries between home and away teams ($t = 1.4565$, $p = 0.1453$). Additionally, there is a significant advantage for the goals scored during the injury time ($t = 4.028$).

Regarding stadium attendance statistics, on average, these top-tier leagues draw around 25000 fans, with peaks nearing 100000 attendees. Given that our dataset also covers the 2019-20 and 2020-21 seasons, we have a subsample of 4542 matches (approximately 12% of

[2] For example, a series of comments from the match Real Madrid-Barcelona on 2014-03-17 includes insights like 'Penalty Barcelona. Neymar draws a foul in the penalty area,' 'Penalty conceded by Sergio Ramos (Real Madrid) after a foul in the penalty area,' and 'Sergio Ramos (Real Madrid) is shown the red card.

[3] Elo scores offer a measure for comparing the overall team strength across short periods.

**Table 2.1:** *Descriptive Statistics*

| Variable | Obs | Mean | SD | Time Frame Min | Time Frame Max | Match Min | Match Max |
|---|---|---|---|---|---|---|---|
| *Time frame variables* | | | | | | | |
| Goals Home | 1800700 | 0.0154 | 0.1232 | 0 | 1 | 0 | 10 |
| Goals Away | 1800700 | 0.0121 | 0.1092 | 0 | 1 | 0 | 9 |
| Fouls committed Home | 1800700 | 0.1214 | 0.3357 | 0 | 4 | 0 | 32 |
| Fouls committed Away | 1800700 | 0.1239 | 0.3390 | 0 | 3 | 0 | 31 |
| Yellow Cards Home | 1800700 | 0.0193 | 0.1386 | 0 | 4 | 0 | 8 |
| Yellow Cards Away | 1800700 | 0.0218 | 0.1472 | 0 | 4 | 0 | 9 |
| Red Cards Home[a] | 1800700 | 0.0009 | 0.0307 | 0 | 3 | 0 | 3 |
| Red Cards Away[a] | 1800700 | 0.0012 | 0.0349 | 0 | 2 | 0 | 3 |
| Penalty Kicks Home | 1800700 | 0.0018 | 0.0423 | 0 | 1 | 0 | 3 |
| Penalty Kicks Away | 1800700 | 0.0013 | 0.0358 | 0 | 1 | 0 | 3 |
| Shots Home | 1800700 | 0.1356 | 0.3832 | 0 | 6 | 0 | 47 |
| Shots Away | 1800700 | 0.1108 | 0.3470 | 0 | 6 | 0 | 38 |
| Injuries Home | 1800700 | 0.0031 | 0.0557 | 0 | 2 | 0 | 3 |
| Injuries Away | 1800700 | 0.0030 | 0.0550 | 0 | 2 | 0 | 4 |
| *Game variables* | | | | | | | |
| Goals extra time Home | 18007 | 0.0819 | 0.2815 | | | 0 | 2 |
| Goals extra time Away | 18007 | 0.0703 | 0.2623 | | | 0 | 2 |
| Injury time 1st | 18007 | 2.1577 | 1.3394 | | | 0 | 16 |
| Injury time 2nd | 18007 | 4.5056 | 1.5765 | | | 0 | 18 |
| Elo Rating | 18007 | 1682.67 | 122.41 | | | 1418 | 2101 |
| Attendance (1000s) | 12219 | 25.43 | 19.51 | | | 0 | 98.76 |
| Closed Doors | 18007 | 0.1262 | - | | | 0 | 1 |

Notes: [a] it is noteworthy that about 40% of the red cards were a result of a player receiving a second yellow card, leading to an automatic red card.

the total) played behind closed doors. We will use this subsample to understand whether there are differences in decisions due to the presence of fans in stadiums.

## 2.3 Methodology

Our investigation into the even-out bias necessitates considering two distinct yet interconnected effects. The first effect concerns the active compensation by referees, potentially driven by a conscious or subconscious desire for fairness and equity. In contrast, the second effect concerns a more passive form of bias, where referees may refrain from making critical calls, especially following a previous non-sanctioned infraction against the opposing team. Both of these scenarios potentially lead to biased application of the rules.

Analytically, if the referees are independent of the past decisions, the following *null hypotheses* must hold:

$$H_0^A : P(C_t = B) = P(C_t = B | C_{<t} = A) \tag{2.1}$$

and

$$H_0^B : P(C_t = B) = P(C_t = B | C_{<t} = B) \tag{2.2}$$

Here, $C$ represents a 'Call', such as awarding a penalty kick (PK) or issuing a red card; A and B represent the two teams involved in the match, without specifying home or away status. The subscript $t$ denotes the time frame, with $t \in (1, T)$, while $< t$ denotes the cumulative time frames before $t$.

Specifically, $H_0^A$ posits that the referee is equally likely to make a call ($C$) against Team B in $t$ if he previously made a similar decision against Team A. Instead, $H_0^B$ suggests that the referee is equally likely to make a call against Team B in $t$ if another similar decision occurred before against Team B.

To analyze these hypotheses, we first assess whether the distribution of these decisions is independent across the matches. We employ the analytical method proposed by Schwarz (2011), which utilizes a Poisson distribution to model the awarding of a penalty or the issuing of a red card. This model relies on an instantaneous event rate, denoted as $\mu(t)$, representing the likelihood of the event happening at any specific time $t$ during a match.

In applying this model, we assume that $\mu(t)$ remains constant throughout a game and is calculated as $\lambda = \frac{N.Events}{Total N.of Matches}$. This approach provides a framework to examine whether the distribution of the event adheres to a pattern that might suggest a deviation from random allocation.

Additionally, we analyze a subset of games where exactly two calls were made by the referee about a certain event. Here a *Chi-squared test* is applied to a *transition matrix* that counts the number of transitions from an event that occurred from the Home team to the Away team, and vice versa, trying to observe any patterns in the distribution of these events. Finally, we conduct an econometric analysis. We opted for a Linear Probability Model (LPM) due to its enhanced interpretability, especially pertinent to the topic under investigation. The LPM's straightforward representation of coefficients as changes in the probability of the dependent event—such as the issuance of yellow cards, red cards, or penalty kicks—makes it particularly suitable for our examination of referees' decision-making processes[4].

Aligning with the methodologies of Scoppa (2021) and Garicano et al. (2005), we consider the point of view of both home and away teams. This entails including each match in the dataset twice: once from the viewpoint of the home team and once from that of the away team. To account for the non-independence of these observations and to ensure robust standard error estimates, during the regression analysis, we will utilize clustered standard errors at the match level.

The main model is as follows:

$$\begin{aligned} Y_{it} = \; & \beta_1 \text{CumSumY\_Team}_{it-1} + \beta_2 \text{CumSumY\_Opponent}_{it-1} \\ & + \beta_3 \text{Home}_i + X_{it}\beta + \alpha_{\text{league}} + \gamma_{\text{season}} \end{aligned}$$

where $Y_{it}$ is a dummy variable that assumes value 1 if in the time frame $t$ of game $i$ a certain event occurred, and 0 otherwise. The variables $EventCount\_Team_{it-1}$ and $EventCount\_Opponent_{it-1}$ represent the count of the respective events for and against both teams up to time frame $t-1$ in-game $i$. $Home_i$ is a dummy variable indicating whether the team is playing at home.

[4] To ensure the robustness of our findings, we conducted an initial analysis using the Logit model, a common alternative that addresses some of the LPM's limitations. This preliminary analysis yielded results that were qualitatively similar to those obtained from the LPM.

The variable $X_{it}$ represents a vector of control variables that account for a range of factors influencing the likelihood of an event occurring in time frame $t$ of game $i$. These factors include the cumulative counts of fouls, cards, penalty kicks, and shots within the game for every time frame $i$. These variables are indicative of the teams' aggressive play and propensity to receive negative or positive calls by referees. To capture the historical context, we incorporate the seasonal rates of penalty kicks and cards, reflecting each team's tendency for such events throughout the season. These rates are the result of the cumulative sum of the events along the season divided by the number of games played at that point of the season. This historical perspective is important for controlling for teams' usual playing styles and their likely impact on the current game.

Additionally, the model accounts for unique match-specific factors, such as the amount of injury time added, the game's attendance, and the ongoing score difference at each time frame, all of which can significantly influence the gameplay and strategic decisions. Furthermore, the Elo scores of both teams are included to control for their overall strength and performance levels, factors that invariably affect the likelihood of various in-game events.

By integrating these diverse variables, $X_{it}$ effectively captures the complex dynamics at play in each match, ranging from immediate in-game strategies to overarching seasonal trends and inherent team characteristics.

## 2.4 Evidence of Even-out Bias of Referees

In our investigation of the even-out bias, we start the analysis from the referees' decisions that could bring the most significant impact on the match, the award of a penalty kick. Subsequently, the focus shifts to red and yellow card decisions, crucial in their respective roles of expelling a player from the game and serving as cautions for rule infringements.

### 2.4.1 Penalty Kicks

The decision to award a penalty kick during a soccer match can be a crucial moment, often significantly influencing the balance of play and potentially determining the match's outcome. The opportunity to take a penalty kick, which has a conversion success rate of approximately 75%, can be highly consequential. Despite the critical nature of such decisions, our data indicates that penalty kicks are not infrequent occurrences; in our dataset comprising 18007 matches, 5530 penalty kicks have been awarded, yielding an average of 0.307 per match, which equates to nearly one penalty every three games.

To assess whether the distribution of these penalty kicks is independent across the matches, we start employing a Poisson distribution to model the awarding of penalty kicks. We assume that $\mu(t)$, the likelihood of a penalty kick being awarded at any specific time $t$, remains constant throughout a game and is calculated as $\lambda = \frac{5530}{18007} \approx 0.307$.

Table 2.2 presents the observed frequencies of matches with 0 to 4 awarded Penalty Kicks in comparison to the expected frequencies as dictated by the Poisson distribution model.

Consistent with the findings of Schwarz (2011), the Poisson distribution appears to be a

**Table 2.2:** *Observed and Predicted number of matches with 0, 1, 2, 3, and 4 Penalty Kicks*

| N. PK | Matches Observed | Matches Predicted | Total PK | Contrib. to $\chi^2$ |
|---|---|---|---|---|
| 0 | 13262 | 13245.50 | 0.00 | 0.02 |
| 1 | 4027 | 4067.73 | 4027.00 | 0.41 |
| 2 | 653 | 624.61 | 1306.00 | 1.29 |
| 3 | 63 | 63.94 | 189.00 | 0.01 |
| 4 | 2 | 4.91 | 8.00 | 1.72 |
| Total | 18007 | 18006.68 | 5530.00 | 3.46 |

*Note:* Predicted frequencies are derived from a Poisson model with $\lambda = \frac{5530}{18007}$.

suitable fit for the distribution of penalties awarded by referees across matches ($\chi^2$(d.f.=3) = 3.4568; $p = 0.1405$). It is observed, however, that there is a lower-than-expected number of matches with a single penalty and a higher number of matches concluding with two penalties. This may be due to an absence of referees' independence from past actions in their decision-making, but this discrepancy alone is not sufficient.

For a first glimpse of the even-out bias, we start analyzing in detail the subsample of matches that have precisely two penalty kicks (Table 2.3). It's important to note that this represents only a subset of the scenarios in which this bias could manifest.

**Table 2.3:** *Transition matrix of the matches with exactly two Penalty Kicks*

| | | Second Penalty Kick | | |
|---|---|---|---|---|
| First Penalty Kick | | **Home** | **Away** | **Total** |
| **Home** | Frq. | 178 | 189 | 367 |
| | Exp. | (209.63) | (157.36) | |
| | Prob. | **0.485** | **0.515** | **0.562** |
| **Away** | Frq. | 195 | 91 | 286 |
| | Exp. | (163.37) | (122.63) | |
| | Prob. | **0.682** | **0.318** | **0.438** |
| Total | | 373 | 280 | 653 |
| Prob. | | **0.571** | **0.429** | |

*Note*: Counts expected under independence are given in brackets.

Our data includes 653 matches where precisely two penalties were awarded, totaling 1306 penalty kicks. Looking at the marginal values of the table, we observe that the home team receives the first penalty in 56.3% of cases, and the second in 57.1%, independent of the team to whom the first penalty kick was awarded. This is in line with the home advantage seen before.

Conversely, taking into account the recipient of the first and second penalty kicks, we notice a dependency. Given the first penalty awarded to the home team, the probability of the home team receiving a second penalty drops to 48.5%. Most notably, when the referee awarded a penalty to the away team, he then awarded a penalty to the home team in 68.2% of matches considered.

Therefore, it seems that referees tend to be dependent on the choices made previously when awarding the second penalty. That is, they tend to reduce the chances of awarding a penalty to the team that has already received one and increase the chances of awarding the second to the rival team. Formally, this assertion is supported by the Chi-square test of independence, which results in $\chi^2 = 25.41$, $p < 0.000$.

Establishing that there is a violation of the independence of referees' decisions concerning penalty awards is not sufficient to ascertain that this is indeed due to a *Even-Out bias*. Multiple factors can influence both referees' decisions and the behavior of the teams after a penalty has been conceded. For instance, as an alternative interpretation, falling behind in the score could lead the losing team to adopt a more offensive strategy, leading to more actions in the opponent penalty area, increasing the likelihood of being awarded a penalty themselves (Mascarenhas et al.; 2002). We delve deeper through a regression analysis.

**Table 2.4:** *Even-out bias of Referees on awarding a Penalty Kick*

| | *Dependent variable:* Penalty Kick in t | | | | | | |
|---|---|---|---|---|---|---|---|
| | (I) All | (II) All | (III) All | (IV) All | (V) All | (VI) Closed Doors | (VII) Closed Doors |
| Constant | 0.002*** | 0.001*** | 0.001*** | 0.008*** | 0.006*** | 0.003*** | 0.008** |
| | (0.0001) | (0.0001) | (0.0001) | (0.001) | (0.001) | (0.0003) | (0.003) |
| **Cumulative P.K.** | **−0.0002*** | **−0.0003**** | **−0.002***** | **−0.004***** | **−0.004***** | **−0.001*** | **−0.004***** |
| | (0.0001) | (0.0001) | (0.0001) | (0.0002) | (0.0002) | (0.0003) | (0.0005) |
| **Cumulative P.K. Opp.** | **0.001***** | **0.001***** | **0.001***** | **0.001***** | **0.001***** | **0.001***** | **0.001**** |
| | (0.0001) | (0.0001) | (0.0001) | (0.0002) | (0.0002) | (0.0004) | (0.0005) |
| Home | | 0.001*** | 0.001*** | 0.001*** | 0.001*** | 0.0003 | 0.0003 |
| | | (0.0001) | (0.0001) | (0.0001) | (0.0001) | (0.0002) | (0.0002) |
| Cumulative Yellow C. | | | −0.0001*** | −0.00005 | −0.00001 | | 0.0001 |
| | | | (0.00004) | (0.00005) | (0.0001) | | (0.0002) |
| Cumulative Yellow C. Opp. | | | 0.001*** | 0.001*** | 0.001*** | | 0.001*** |
| | | | (0.0001) | (0.0001) | (0.0001) | | (0.0002) |
| Cumulative Red C. | | | −0.001*** | −0.001*** | −0.001*** | | −0.001** |
| | | | (0.0002) | (0.0002) | (0.0002) | | (0.001) |
| Cumulative Red C. Opp. | | | 0.006*** | 0.006*** | 0.005*** | | 0.005*** |
| | | | (0.0003) | (0.0003) | (0.0004) | | (0.001) |
| Cumulative Fouls | | | −0.0001*** | −0.0001*** | −0.0001*** | | −0.0001*** |
| | | | (0.00001) | (0.00001) | (0.00001) | | (0.00004) |
| Cumulative Fouls Opp. | | | 0.0002*** | 0.0002*** | 0.0003*** | | 0.0004*** |
| | | | (0.00001) | (0.00001) | (0.00002) | | (0.00005) |
| Cumulative Shots | | | 0.00001 | −0.00003** | −0.00001 | | −0.0001** |
| | | | (0.00001) | (0.00001) | (0.00001) | | (0.00004) |
| Cumulative Shots Opp. | | | −0.0001*** | −0.0001*** | −0.0001*** | | −0.0001*** |
| | | | (0.00001) | (0.00001) | (0.00001) | | (0.00003) |
| P.K. rate season | | | | 0.0009*** | 0.001*** | | 0.001*** |
| | | | | (0.0002) | (0.0002) | | (0.0002) |
| P.K. rate season Opp. | | | | −0.0000 | −0.0000 | | −0.0000 |
| | | | | (0.0002) | (0.0002) | | (0.0001) |
| Difference in Score | | | | | 0.0002*** | | 0.0002** |
| | | | | | (0.00004) | | (0.0001) |
| Injury Time 1H | | | | | 0.0002*** | | 0.0003*** |
| | | | | | (0.00003) | | (0.0001) |
| Injury Time 2H | | | | | 0.0001*** | | 0.0001* |
| | | | | | (0.00003) | | (0.0001) |
| League and Season FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Other Seasonal Rate vars. | No | No | No | Yes | Yes | No | Yes |
| Game-Specific variables | No | No | No | No | Yes | No | Yes |
| Observations | 3,601,400 | 3,601,400 | 3,601,400 | 3,021,000 | 3,021,000 | 454,200 | 454,200 |
| Adjusted $R^2$ | 0.0001 | 0.0001 | 0.001 | 0.004 | 0.004 | 0.004 | 0.004 |

*Note:* The analysis employs a linear probability model. Standard errors are robust to heteroskedasticity and are clustered at the game level. The other seasonal rate variables include the rates for both teams of red cards and yellow cards. The game-specific variables are a dummy for the half of the match, the attendance, and the Elo scores. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 2.4 (columns I-VII) presents the outcomes of multiple linear probability models, exploring the likelihood of awarding penalty kicks within distinct time frames $t$. The coefficients' small size across all regressions is due to the division of the game into 100 frames, which naturally lowers the likelihood of a penalty kick in any single frame. This division significantly reduces the probability of observing a penalty kick in any singular frame, yet it facilitates the exploration of the dynamics guiding referees' adjudication processes on a moment-to-moment basis.

The coefficients for *Cumulative P.K.* and *Cumulative P.K. Opponent* shed light on referee behavior regarding penalty kicks. The negative coefficient for Cumulative P.K. (-0.0002, $p < 0.1$) suggests referees tend to avoid awarding successive penalty kicks to the same team. In contrast, the positive coefficient for Cumulative P.K. Opponent (0.001, $p < 0.01$) indicates a higher likelihood of a team receiving a penalty kick if the opposing team has previously received one. These findings contradict both $H_0^A$ and $H_0^B$, pointing to a subconscious effort by referees to maintain fairness.

Column (II) introduces control for potential home advantage. The significant positive coefficient for Home (0.001, $p < 0.01$) suggests a subtle home advantage in referee decisions, aligning with existing literature on the subject (Dohmen; 2008; Dohmen and Sauermann; 2016; Scoppa; 2021).

Columns (III) and (IV) extend the analysis by controlling for game aggressiveness, intensity, and seasonal trends. The coefficients for *Cumulative Red C.* and *Cumulative Red C. Opponent* provide a deeper understanding of how referees' past decisions on red cards influence their subsequent penalty kick decisions. Instead, the positive and statistically significant coefficient of *P.K. rate season* suggests that teams receiving more penalties in the past matches are more likely to be awarded one in this match; this could be an effect of tactical style or players' skills. Column (V) presents the results for the model with the full set of controls; as expected, an increase in the injury time is associated with a higher chance of being awarded a penalty kick.

The models in columns VI and VII, focus exclusively on games played behind closed doors. This allows for a comparison of the crowd presence on referee behavior. Interestingly, the coefficient for the variable *Home* is not significant, indicating that the absence of a crowd diminishes the home team's advantage. A key observation across all the regressions is the consistent significance of the coefficients related to the even-out bias, *Cumulative P.K.* and *Cumulative P.K. Opponent*.

To better understand the effects of previous decisions on the current probability of assigning a penalty kick, we can express the coefficients in terms of change in the baseline expected probability of seeing a penalty kick in that time frame, 0.0033, or 0.33%.

Taking the results of the model in column V, the effect of *Cumulative P.K*, the cumulative sum of penalty kicks for the team, relative to the baseline penalty kick probability per time frame, is $\frac{-0.004}{0.0033} = -1.21$. This ratio indicates that each additional penalty kick awarded to a team reduces the chance of receiving another by 121% compared to the base probability of a penalty award within any minute. This significant decrease, while reflecting a modest absolute change due to the inherent rarity of penalty kicks, points to a substantial relative reduction.

Conversely, the effect of a penalty awarded to the opposing team, captured as $\frac{0.002}{0.0033} = 0.303$, indicates that having awarded a penalty to the opponent increases the probability of the understudy team receiving a penalty by 30%, relative to the baseline probability. These observations point to referees' subconscious efforts to mitigate bias, potentially reflecting an intrinsic motivation towards ensuring fairness within the game.

These findings support the hypothesis that referees are influenced by past events in their decision-making process, striving to balance their calls in a manner that compensates for earlier decisions. Such behavior underscores the complex dynamics of fairness and bias in sports officiating, offering valuable insights into the psychological factors influencing referees' decisions in high-stakes environments.

### 2.4.2 Red Cards

Our focus now shifts to exploring the presence of the even-out bias in the allocation of red cards. Similar to penalty kick decisions, issuing a red card can significantly alter the game balance and the outcome of a soccer match. This action mandates the offending player's departure from the field, consequently reducing their team to a numerical disadvantage. For this reason, the sanction of sending off a player is undertaken in specific circumstances dictated by the Law of the Game, IFAB (2023). Among the many offenses punishable by a direct red card are the player's violent or offensive conduct, denial of a goal or an obvious goal-scoring opportunity with an offense, or serious foul play that endangers the safety of an opponent or uses excessive force or brutality (Law 12, Laws of the Game IFAB (2023)). Moreover, a red card is issued to the player who received two yellow cards during the same match.

Following the approach seen before on the decision of penalty kicks, our initial step involves examining whether the issuance of successive red cards within a single game is independent of prior similar events.

Considering the infrequency of red cards in soccer, with an average occurrence of one in every five games, we adopt a Poisson distribution as our analytical model, focusing on the parameter lambda. To ensure the accuracy of this analysis, we chose to exclude three games in which the red cards were issued post-match[5].

In our sample of 18004 games, referees issued a total of 3846 red cards, which translates to an average of approximately 0.21 red cards per match. Consistent with findings of home favoritism and home advantages (Boyko et al.; 2007; Dohmen and Sauermann; 2016; Scoppa; 2021), a significant proportion of these red cards, around 60%, were shown by the referees to the away team.

As seen before, during any given moment $t$ in a match, there exists an instantaneous event rate $\mu(t)$, representing the likelihood of a red card being issued at that specific point. Table 2.5 shows a comparative display of the observed number of matches with varying red card occurrences against the theoretical frequencies predicted by the Poisson distribution model. The analysis reveals two significant deviations from the expected Poisson frequencies ($\chi^2$ (d.f.= 4) = of 58.67, $p$<0.000). The first deviation is the lower-than-expected frequency of matches with a single red card, while there is a noticeable increase in matches where exactly two or three red cards were issued.

The observation of a deviation in the distribution of red cards across the matches is an initial finding that warrants further exploration, but it is not conclusive. Table 2.6 presents the transition matrix of the 396 matches where exactly two red cards were issued by the referees.

[5] Referees have the authority to take disciplinary actions until leaving the field of play after the match ends IFAB (2023).

**Table 2.5:** *Observed and Predicted number of matches with 0, 1, 2, 3, and 4 red cards*

| N.Red Cards | Matches Observed | Matches Predicted | Total Red Cards | Contribution to $\chi^2$ |
|---|---|---|---|---|
| 0 | 14669 | 14529.45 | 0.00 | 1.34 |
| 1 | 2872 | 3115.35 | 2872.00 | 19.01 |
| 2 | 419 | 333.99 | 838.00 | 21.64 |
| 3 | 40 | 23.87 | 120.00 | 10.90 |
| 4 | 4 | 1.28 | 16.00 | 5.78 |
| Total | 18004 | 18004 | 3846 | 58.67 |

Notes: Predicted frequencies are derived from the Poisson model with $\lambda = 3846/18004 = 0.2144$.

exclude from our analysis the games where the referee issued a red card simultaneously to players of both teams, as well as games in which the red cards were issued post-match.

**Table 2.6:** *Transition matrix of the matches with exactly two Red Cards*

| | | Second Red Card | | |
|---|---|---|---|---|
| First Red Card | | **Home** | **Away** | **Total** |
| | Frq. | 69 | 112 | 181 |
| **Home** | Exp. | (83.7) | (97.4) | |
| | Prob. | **0.381** | **0.619** | **0.457** |
| | Frq. | 114 | 101 | 215 |
| **Away** | Exp. | ( 99.4) | (115.6) | |
| | Prob. | **0.53** | **0.46** | **0.543** |
| Total | | 183 | 213 | 396 |
| **Prob.** | | **0.462** | **0.538** | |

Notes: Counts expected under independence are given in brackets.

Distinctively, there is a pronounced occurrence of matches where both participating teams received one red card each, as indicated by the numbers on the main diagonal. In detail, among the matches in which exactly two red cards were recorded, the home team received the first card and the second card in 45.7% and 46.2% of the matches with exactly two red cards, respectively. Looking instead at the conditional probabilities, when the home team received the first card, the frequency in which it also received the second card dropped to 38.1%, while, given the card to the away team, the probability of the second card to the home team they rise to 53%. The probabilities shown in the table are in contrast with the null hypothesis of independent decisions taken by the referees (Equations. 2.1 and 2.2). This dependence is statistically significant ($\chi^2$ = 8.779, p = 0.003).

This analysis, however, included only a subsample of cases in which referees can express this dependence on past actions. Furthermore, as with penalty kicks, it is essential to take into account the dynamics present during matches in a context such as soccer. For this reason, we extend our investigation to include regression analysis.

Table 2.7 presents the outcomes of the regression analysis, focusing on the probability of being issued a red card by referees during the time frame $t$. Consistent with prior exclusions, instances where red cards are shown after the game are omitted from the analysis. Columns I and II report the simple model and the introduction of the variable *Home*, respectively. Initially, we notice a positive effect of having received a red card previously on the possibility of receiving it in time t. This would contradict what we saw previously for penalty kicks. Instead, the coefficient relating to the cumulative sum of the opponent's reds is in line

**Table 2.7:** *Even-out bias of Referees on issuing a Red Card*

| | *Dependent variable:* Red Card in t | | | | | | |
|---|---|---|---|---|---|---|---|
| | (I) All | (II) All | (III) All | (IV) All | (V) All | (VI) Closed Doors | (VII) Closed Doors |
| Constant | 0.001*** | 0.002*** | 0.0003** | −0.002*** | −0.004*** | 0.001*** | −0.006 |
| | (0.0001) | (0.0001) | (0.0001) | (0.001) | (0.001) | (0.0002) | (0.004) |
| **Cumulative Red C.** | **0.001***** | **0.001***** | **−0.001***** | **−0.003***** | **−0.003***** | **0.001**** | **−0.002***** |
| | (0.0002) | (0.0002) | (0.0002) | (0.0003) | (0.0003) | (0.001) | (0.001) |
| **Cumulative Red C. Opp.** | **0.002***** | **0.002***** | **0.001***** | **0.001***** | **0.001***** | **0.003***** | **0.002*** |
| | (0.0002) | (0.0002) | (0.0002) | (0.0002) | (0.0003) | (0.001) | (0.001) |
| Home | | −0.0004*** | −0.0003*** | −0.0003*** | −0.0003*** | −0.0003* | −0.0003* |
| | | (0.0001) | (0.0001) | (0.0001) | (0.0001) | (0.0001) | (0.0002) |
| Cumulative Yellow C. | | | 0.001*** | 0.001*** | 0.001*** | | 0.001*** |
| | | | (0.00005) | (0.0001) | (0.0001) | | (0.0002) |
| Cumulative Yellow C. Opp. | | | 0.0002*** | 0.0002*** | 0.0002*** | | 0.0004*** |
| | | | (0.00005) | (0.00005) | (0.0001) | | (0.0001) |
| Cumulative P.K. | | | −0.00002 | 0.00000 | −0.00003 | | 0.00002 |
| | | | (0.0001) | (0.0001) | (0.0002) | | (0.0003) |
| Cumulative P.K. Opp. | | | 0.0003** | 0.0004*** | 0.0003* | | 0.0005 |
| | | | (0.0002) | (0.0002) | (0.0002) | | (0.0004) |
| Cumulative Fouls | | | 0.0001*** | 0.0001*** | 0.0001*** | | 0.0001** |
| | | | (0.00001) | (0.00001) | (0.00001) | | (0.00004) |
| Cumulative Fouls Opp. | | | −0.00001 | −0.00002* | −0.00001 | | −0.00003 |
| | | | (0.00001) | (0.00001) | (0.00001) | | (0.00004) |
| Cumulative Shots | | | −0.00002*** | −0.00002* | −0.00000 | | −0.00001 |
| | | | (0.00001) | (0.00001) | (0.00001) | | (0.00003) |
| Cumulative Shots Opp. | | | 0.00001 | 0.00001 | 0.00003** | | 0.00001 |
| | | | (0.00001) | (0.00001) | (0.00001) | | (0.00003) |
| R.C. rate season | | | | −0.00002 | −0.0001 | | −0.0001 |
| | | | | (0.0002) | (0.0002) | | (0.0002) |
| R.C. rate season Opp. | | | | −0.0003 | −0.0003 | | −0.0002 |
| | | | | (0.0003) | (0.0002) | | (0.0002) |
| Difference in Score | | | | | −0.0001* | | −0.00002 |
| | | | | | (0.00003) | | (0.0001) |
| Injury Time 1H | | | | | 0.0002*** | | 0.0001* |
| | | | | | (0.00003) | | (0.0001) |
| Injury Time 2H | | | | | 0.0002*** | | 0.0002*** |
| | | | | | (0.00003) | | (0.0001) |
| League and Season FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Other Seasonal Rate vars. | No | No | No | Yes | Yes | No | Yes |
| Game-Specific variables | No | No | No | No | Yes | No | Yes |
| Observations | 3,600,800 | 3,600,800 | 3,600,800 | 3,600,800 | 3,020,600 | 454,200 | 454,200 |
| $R^2$ | 0.0002 | 0.0003 | 0.002 | 0.004 | 0.004 | 0.0003 | 0.004 |

*Note:* The analysis employs a linear probability model. Standard errors are robust to heteroskedasticity and are clustered at the game level. The other seasonal rate variables include the rates for both teams of penalty kicks and yellow cards. The game-specific variables are a dummy for the half of the match, the attendance, and the Elo scores. *p<0.1; **p<0.05; ***p<0.01

with the hypothesis that referees are affected by even-out bias. Again, as seen for penalties, there is an advantage of playing home, as fewer red cards are received on average (-0.0004, p<0.001).

By incorporating variables that account for match intensity, including the number of fouls, shots, and previously awarded penalties, we observe a notable shift in the *Cumulative R.C.* variable's sign. Specifically, the introduction of these intensity-related variables in our model reveals a consistent pattern where referees adjust their decisions to balance previous disciplinary actions. Contrary to the observations made for penalty kicks, there is no discernible correlation between the rate of red cards received during the season and the likelihood of receiving a red card in the current match. (column IV). Finally, analyzing the context of the matches played without spectators (columns VI and VII), we note similar results to what we saw previously; the introduction of control variables changes the sign of the *Cumulative R.C.* coefficient, which is significant in both cases.

Building on the approach used to analyze penalty kicks, we apply a similar methodology to understand the influence of past red card decisions on the likelihood of future red card issuance. Utilizing the baseline probability for a red card occurrence within any given time frame, set at 0.214%, we examine the coefficients from Column V to quantify these effects. For the *Cumulative Red C.* variable, representing the cumulative sum of red cards issued against the team, the calculation $\frac{-0.003}{0.0214} = -1.39$ signifies a pronounced decrease in the probability of a team receiving another red card, by 139% relative to the baseline, for each additional red card previously allocated. This substantial relative reduction mirrors our findings with penalty kicks, highlighting a consistent pattern of decision adjustment by referees to avoid over-penalization and maintain fairness.

Similarly, analyzing the impact of red cards given to the opponent (*Cumulative Red C. Opponent*), we find an effect of $\frac{0.001}{0.00214} = 0.4664$, translating to a 46.64% increase in the chance of a team receiving a red card following one awarded to their opponent.

These analyses affirm that referees' decisions on red cards are influenced by previous events, in a manner akin to their adjudications on penalty kicks. Such a strategy of compensatory behavior indicates referees' underlying efforts to ensure fairness, reflecting the intricate interplay of past decisions, game dynamics, and the psychological factors at play in high-stakes officiating situations.

### 2.4.3 Yellow Cards

Following our detailed examination of red cards for potential even-out bias, we now extend our analysis to another critical aspect of referee decision-making: yellow card issuance. According to the Laws of the Game (IFAB; 2023), referees use yellow cards to sanction a series of offenses or unsporting behavior by players. For some offenses, the regulation leaves no room for interpretation, such as removing the shirt during the goal celebration. But there are circumstances in which the use of this sanction depends on the referee's interpretation such as the severity of the fouls committed or non-serious unsporting behavior.

The first evidence of the even-out bias is presented in Fig. 2.1. The plot depicts the probability that the next yellow card will be issued against the home team, based on the net difference in cumulative yellow cards before the yellow was issued. Notably, when the home team has received fewer yellow cards (negative net difference), there is a significantly higher probability of them receiving the next yellow card. This probability decreases as the net difference becomes less negative or turns positive, suggesting a tendency among referees to 'balance out' the distribution of yellow cards.

This trend, statistically significant, (F=31.5, df=10, p<0.000), aligns with the theory of referees, consciously or subconsciously, seeking fairness in their decisions.

Table 2.8 further substantiates the even-out bias in referees' decision-making process regarding the issuance of yellow cards. The regression models across Columns I-VII in the table systematically incorporate varying game dynamics, aggressiveness, and team history, painting a comprehensive picture of the factors influencing referees' decisions. Notably, the coefficients for *Cumulative P.K.* and *Cumulative P.K. Opponent* are indicative of the impact past events within the match have on subsequent rulings. We can reject both $H_0^A$ and $H_0^B$.

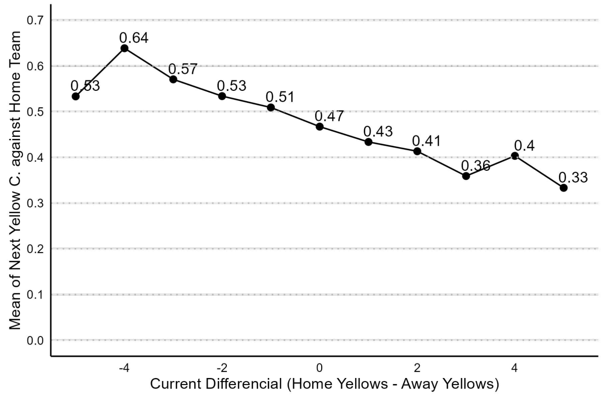


**Figure 2.1:** *Mean of Next Yellow C. against Home Team by Current Net Differential*

Notably, the negative coefficient for *Cumulative P.K.* across all the columns suggests a reluctance to continuously penalize the same team, reflecting an effort to maintain an equitable distribution of penalties.

In Table 2.8, the regression analysis delves into the factors affecting the issuance of yellow cards across our sample of soccer matches. Columns I and II serve as the foundational models, where the coefficients for *Cumulative P.K.* and *Cumulative P.K. Opponent* initially indicate the immediate impact of penalty kicks on the likelihood of yellow cards being issued. The negative coefficient for *Cumulative P.K.* in some columns suggests a slight tendency to avoid penalizing the same team repeatedly within a short time frame, while the positive coefficient for *Cumulative P.K. Opponent* indicates an increased likelihood of issuing yellow cards to the opposing team. Furthermore, as indicated by the positive coefficients of the seasonal rates, an increase in the frequency with which referees penalized both teams correlates with a higher probability of imposing sanctions in the current match.

Interestingly, in games without crowds (Columns VI and VII), the usual home-field advantage, typically reflected in referee decisions, appears to be diminished. This observation points to the significant influence of external factors, such as crowd presence, on referee behavior in issuing yellow cards.

Considering the overall frequency of yellow cards at about 4.11 per game, the issuance of a yellow card is relatively common compared to penalty kicks and red cards. Reflecting on this higher baseline occurrence rate, we examine the relative effects of past yellow card decisions on the probability of a yellow card being issued within the current time frame $t$. The coefficient for *Cumulative Yellow C.* from Column V, indicating the cumulative sum of yellow cards given to the team, reveals a decrease in the chance of a team receiving another yellow card by 21.89%, calculated as $\frac{-0.009}{0.0411} = -0.2189$. This reduction, though not as pronounced as for red cards, indicates a careful approach by referees to avoid excessively penalizing teams. On the other hand, the coefficient for yellow cards awarded to the opponent (*Cumulative Yellow C. Opponent*), at 0.003, results in a 7.3% increase in the likelihood of a team being issued a yellow card, computed from $\frac{0.003}{0.0411} = 0.073$. This effect suggests that referees are likely adjusting their decisions to ensure disciplinary actions are distributed more evenly between teams, further demonstrating their effort to maintain fairness within the match.

**Table 2.8:** *Even-out bias of Referees on issuing a Yellow Card*

| | *Dependent variable:* Yellow Card in t | | | | | | |
|---|---|---|---|---|---|---|---|
| | (I) All | (II) All | (III) All | (IV) All | (V) All | (VI) Closed Doors | (VII) Closed Doors |
| Constant | 0.023*** | 0.026*** | 0.014*** | −0.031*** | −0.054*** | 0.023*** | −0.056*** |
| | (0.0004) | (0.0004) | (0.0005) | (0.003) | (0.004) | (0.001) | (0.009) |
| **Cumulative Yellow C.** | **−0.0001** | **−0.0003**** | **−0.007***** | **−0.009***** | **−0.009***** | **−0.00004** | **−0.008***** |
| | (0.0001) | (0.0001) | (0.0002) | (0.0002) | (0.0002) | (0.0004) | (0.001) |
| **Cumulative Yellow C. Opp.** | **0.007***** | **0.007***** | **0.003***** | **0.004***** | **0.003***** | **0.007***** | **0.003***** |
| | (0.0001) | (0.0001) | (0.0002) | (0.0002) | (0.0002) | (0.0004) | (0.001) |
| Home | | −0.005*** | −0.004*** | −0.005*** | −0.005*** | −0.0002 | −0.0002 |
| | | (0.0002) | (0.0002) | (0.0003) | (0.0003) | (0.001) | (0.001) |
| Cumulative Red C. | | | −0.003*** | −0.001* | −0.002*** | | −0.002 |
| | | | (0.001) | (0.001) | (0.001) | | (0.002) |
| Cumulative Red C. Opp. | | | 0.004*** | 0.004*** | 0.004*** | | 0.005** |
| | | | (0.001) | (0.001) | (0.001) | | (0.002) |
| Cumulative P.K. | | | 0.0005 | 0.001** | 0.001* | | 0.003** |
| | | | (0.001) | (0.001) | (0.001) | | (0.001) |
| Cumulative P.K. Opp. | | | −0.001 | −0.0003 | −0.0004 | | 0.001 |
| | | | (0.001) | (0.001) | (0.001) | | (0.001) |
| Cumulative Fouls | | | 0.002*** | 0.002*** | 0.002*** | | 0.002*** |
| | | | (0.0001) | (0.0001) | (0.0001) | | (0.0002) |
| Cumulative Fouls Opp. | | | 0.0003*** | 0.0004*** | 0.0005*** | | 0.001*** |
| | | | (0.00005) | (0.0001) | (0.0001) | | (0.0002) |
| Cumulative Shots | | | −0.0001*** | 0.00004 | 0.0001** | | 0.0001 |
| | | | (0.00003) | (0.00004) | (0.00004) | | (0.0001) |
| Cumulative Shots Opp. | | | 0.001*** | 0.0004*** | 0.001*** | | 0.001*** |
| | | | (0.00004) | (0.00004) | (0.00004) | | (0.0001) |
| Y.C. rate season | | | | 0.003*** | 0.003*** | | 0.003*** |
| | | | | (0.0003) | (0.0002) | | (0.001) |
| Y.C. rate season Opp. | | | | 0.0007** | 0.0008** | | 0.0008*** |
| | | | | (0.0003) | (0.0002) | | (0.0002) |
| Injury Time 1H | | | | | 0.001*** | | 0.001*** |
| | | | | | (0.0001) | | (0.0003) |
| Injury Time 2H | | | | | 0.002*** | | 0.002*** |
| | | | | | (0.0001) | | (0.0002) |
| League and Season FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Other Seasonal Rate vars. | No | No | No | Yes | Yes | No | Yes |
| Game-Specific variables | No | No | No | No | Yes | No | Yes |
| Observations | 3,601,400 | 3,601,400 | 3,601,400 | 3,601,400 | 3,021,000 | 454,200 | 454,200 |
| Adjusted $R^2$ | 0.0002 | 0.0003 | 0.002 | 0.004 | 0.004 | 0.002 | 0.007 |

*Note:* The analysis employs a linear probability model. Standard errors are robust to heteroskedasticity and are clustered at the game level. The other seasonal rate variables include the rates for both teams of penalty kicks and red cards. The game-specific variables are a dummy for the half of the match, the attendance, and the Elo scores. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Overall, the findings across these models resonate with the idea that referees, perhaps unconsciously, strive to balance their decisions in an attempt to uphold fairness. This tendency, while subtle, is a critical aspect of maintaining the integrity and fairness of the sport.

### 2.4.4 Methodological Checks

As a methodological check for testing the stability of our results, we conducted an additional regression that involves segmenting the matches into $T = 10$ time frames instead of $T = 100$. The results are shown in Table 2.9. For each of the refereeing decisions analyzed previously, we report the simple model with only the home advantage and even-out bias variables, and the complete model with all the controls seen previously. As expected, the overall Adjusted $R^2$ for these regressions is higher compared to the scenario with $T = 100$ time frames, indicating a stronger explanatory power of our models when games are segmented

**Table 2.9:** *Robustness check on the Even-out bias*

| | *Dependent variable:* | | | | | |
|---|---|---|---|---|---|---|
| | Penalty Kick | | Red Card | | Yellow Card | |
| | (I) | (II) | (III) | (IV) | (V) | (VI) |
| Constant | 0.008$^{***}$ | −0.010$^{*}$ | 0.010$^{***}$ | −0.012$^{**}$ | 0.153$^{***}$ | −0.150$^{***}$ |
| | (0.001) | (0.006) | (0.001) | (0.005) | (0.003) | (0.022) |
| Home | 0.005$^{***}$ | 0.003$^{***}$ | −0.003$^{***}$ | −0.001$^{***}$ | −0.029$^{***}$ | −0.024$^{***}$ |
| | (0.0004) | (0.0005) | (0.0003) | (0.0004) | (0.002) | (0.002) |
| **Cumulative P.K.** | **−0.001** | **−0.011**$^{***}$ | | 0.001 | | 0.003 |
| | (0.001) | (0.001) | | (0.001) | | (0.004) |
| **Cumulative P.K. Opp.** | **0.006**$^{***}$ | **0.009**$^{***}$ | | −0.0001 | | −0.019$^{***}$ |
| | (0.001) | (0.001) | | (0.001) | | (0.004) |
| **Cumulative R.C.** | | −0.001 | **0.005**$^{***}$ | **−0.005**$^{***}$ | | −0.035$^{***}$ |
| | | (0.001) | (0.001) | (0.002) | | (0.006) |
| **Cumulative R.C. Opp.** | | 0.006$^{***}$ | **0.013**$^{***}$ | **0.005**$^{***}$ | | 0.022$^{***}$ |
| | | (0.002) | (0.002) | (0.002) | | (0.006) |
| **Cumulative Y.C.** | | 0.001$^{***}$ | | 0.007$^{***}$ | **0.007**$^{***}$ | **−0.038**$^{***}$ |
| | | (0.0003) | | (0.0004) | (0.001) | (0.001) |
| **Cumulative Y.C. Opp.** | | −0.0002 | | 0.001$^{***}$ | **0.050**$^{***}$ | **0.017**$^{***}$ |
| | | (0.0004) | | (0.0003) | (0.001) | (0.001) |
| Cumulative Fauls | | −0.0002 | | 0.001$^{***}$ | | 0.014$^{***}$ |
| | | (0.0001) | | (0.0001) | | (0.0004) |
| Cumulative Fauls Opp. | | 0.001$^{***}$ | | 0.0002 | | 0.005$^{***}$ |
| | | (0.0001) | | (0.0001) | | (0.0004) |
| Cumulative Shots | | 0.0001 | | 0.0002$^{***}$ | | 0.002$^{***}$ |
| | | (0.0001) | | (0.0001) | | (0.0003) |
| Cumulative Shots Opp. | | 0.0004$^{***}$ | | 0.0001 | | 0.004$^{***}$ |
| | | (0.0001) | | (0.0001) | | (0.0003) |
| P.K. rate season | | 0.007$^{***}$ | | 0.001 | | −0.007 |
| | | (0.002) | | (0.001) | | (0.006) |
| P.K. rate season Opp. | | −0.0004 | | −0.001 | | −0.012$^{**}$ |
| | | (0.002) | | (0.001) | | (0.006) |
| R.C. rate season | | −0.003 | | −0.001 | | −0.0002 |
| | | (0.002) | | (0.002) | | (0.008) |
| R.C. rate season Opp. | | −0.0004 | | −0.002 | | 0.010 |
| | | (0.002) | | (0.002) | | (0.008) |
| Y.C. rate season | | 0.001 | | 0.001$^{***}$ | | 0.019$^{***}$ |
| | | (0.0005) | | (0.0004) | | (0.002) |
| Y.C. rate season Opp. | | −0.0004 | | −0.0002 | | 0.006$^{***}$ |
| | | (0.001) | | (0.0004) | | (0.002) |
| Score Diff. | | 0.007$^{***}$ | | −0.002$^{***}$ | | −0.008$^{***}$ |
| | | (0.0003) | | (0.0002) | | (0.001) |
| Injury Time 1H | | 0.002$^{***}$ | | 0.001$^{***}$ | | 0.007$^{***}$ |
| | | (0.0002) | | (0.0002) | | (0.001) |
| Injury Time 2H | | 0.001$^{***}$ | | 0.001$^{***}$ | | 0.014$^{***}$ |
| | | (0.0002) | | (0.0002) | | (0.001) |
| League and Season FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Other Seasonal Rate vars. | No | Yes | No | Yes | No | Yes |
| Game-Specific variables | No | Yes | No | Yes | No | Yes |
| Observations | 360,140 | 293,380 | 360,140 | 293,380 | 360,140 | 293,380 |
| Adjusted $R^2$ | 0.001 | 0.008 | 0.002 | 0.011 | 0.020 | 0.043 |

*Note:* The analysis employs a linear probability model. Standard errors are robust to heteroskedasticity and are clustered at the game level. The game-specific variables are a dummy for the half of the match, the attendance, and the Elo scores. $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01

into broader intervals. In general, the effects seen so far are confirmed even if the matches are divided into 10 frames. This suggests that referees' propensity towards an even-out bias persists even for longer segments.

## 2.5 Discussion

This study's exploration into the compensatory biases of referees in professional soccer has yielded significant insights. Our findings reveal that referees, possibly driven by an inherent desire for fairness and balance, exhibit a tendency to 'even out' decisions across teams. This is particularly evident in scenarios involving penalty kicks, red, and yellow cards.

These tendencies align with the broader literature on behavioral economics and sports psychology, which suggests that non-monetary incentives such as social pressure, the pursuit of fairness, and the desire to avoid criticism can significantly influence decision-making. The observed even-out bias suggests that referees may subconsciously aim to balance their calls to prevent perceived inequity, which could influence the outcome of matches. Our findings are consistent with those of Plessner and Betsch (2001) where referees showed an influence of previous decisions in similar situations in a controlled experimental setting, where there is no influence of the crowd or players on the field. The presence of an even-out bias challenges the traditional view of referees as purely objective arbiters and underscores the complex psychological factors that play a role in their decision-making process. It also resonates with theories from behavioral economics, particularly those proposed by Fehr and Schmidt (1999), about the concept of inequity aversion. Individuals prefer fair outcomes and may adjust their behavior to achieve perceived equity.

However, alternative interpretations are possible. Rather than a pure aversion to inequity, what might drive referees is a sense of atonement. They may try to compensate for their previous unfavorable calls by making a subsequent favorable call for the other team. This perspective shifts the focus from a desire for fairness to a mechanism of rectifying perceived past mistakes. Additionally, experienced players and coaches might leverage previous mistakes or calls by referees to sway their subsequent decisions. This form of lobbying and complaining by players on the field after a decision could also play a role in the phenomenon of balancing previous calls.

Contextualizing these findings within the broader spectrum of decision-making and behavioral economics, it becomes clear that the environment in which referees operate is replete with factors that can sway their judgment. The role of crowd noise, media scrutiny, and the high stakes associated with professional sports can create an intense environment, where the desire to maintain fairness and impartiality becomes a critical driving force.

Moreover, this study's implications extend beyond the soccer field, offering valuable insights into human behavior in high-pressure decision-making scenarios. Similar biases and tendencies may be observable in other professions where impartiality and fairness are paramount. For instance, judges in legal courts, umpires in other sports, or even evaluators in academic and corporate settings might exhibit similar compensatory behaviors in their decision-making processes.

Given these findings, the implementation of new technologies, such as VAR (Video Assistant Referee), becomes increasingly important. Such technologies can help address biases in refereeing decisions, ensuring that the game's outcomes are determined more by the players' skills and strategies rather than the subjective judgments of referees. This technological intervention could be a crucial step towards ensuring real fairness and objectivity in professional soccer.

# References

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

Boyko, R. H., Boyko, A. R. and Boyko, M. G. (2007). Referee bias contributes to home advantage in english premiership football, **25**(11): 1185–1194.

Dohmen, T. J. (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, **30**(4): 679–695.

Fehr, E. and Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation*, *The Quarterly Journal of Economics* **114**(3): 817–868.

Garicano, L., Palacios-Huerta, I. and Prendergast, C. (2005). Favoritism under social pressure, **87**(2): 208–216.

Haynes, G. and Gilovich, T. (2010). "the ball don't lie": How inequity aversion can undermine performance, **46**(6): 1148–1150.

IFAB (2023). *Laws of the Game*, International Football Association Board.
**URL:** *https://www.theifab.com/laws*

Mascarenhas, D., 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.

Noecker, C. A. and Roback, P. (2012). New insights on the tendency of NCAA basketball officials to even out foul calls, **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* **23**(3): 254 – 259.

Robberechts, P., Van Haaren, J. and Davis, J. (2019). Who will win it? an in-game win probability model for football.

Schwarz, W. (2011). Compensating tendencies in penalty kick decisions of referees in professional football: Evidence from the german bundesliga 1963–2006, **29**(5): 441–447.

Scoppa, V. (2021). Social pressure in the stadiums: Do agents change behavior without crowd support?, **82**.

Wertheim, L. J. (2011). *Scorecasting*, Crown/Archetype.

# Appendix

**Table B1:** *Number of Games per season across Top 5 European Leagues*

| Season | Serie A | Ligue 1 | Premier League | Bundes-liga | La Liga | Total |
|---|---|---|---|---|---|---|
| 12-13 | 379 | 365 | 294 | 306 | 380 | 1724 |
| 13-14 | 377 | 379 | 377 | 293 | 380 | 1806 |
| 14-15 | 380 | 378 | 377 | 306 | 380 | 1821 |
| 15-16 | 379 | 376 | 373 | 306 | 379 | 1813 |
| 16-17 | 380 | 379 | 380 | 305 | 380 | 1824 |
| 17-18 | 380 | 379 | 380 | 306 | 380 | 1825 |
| 18-19 | 380 | 380 | 380 | 306 | 380 | 1826 |
| 19-20 | 379 | 279 | 380 | 306 | 380 | 1724 |
| 20-21 | 380 | 380 | 379 | 306 | 375 | 1820 |
| 21-22 | 380 | 380 | 378 | 306 | 380 | 1824 |

There are some missing matches from the 2012-13 Premier League due to the different format and structure of the available commentary. Instead, the absence of some games during the 2019-20 season in Ligue 1 is attributed to the premature termination of the championship due to the COVID-19 pandemic.

# Chapter 3

# Home advantage and Gender Differences: Evidence from Major Women's European leagues

## Abstract

This study delves into the home advantage phenomenon in major women's soccer leagues, with a novel focus on crowd density—the ratio of spectators to available seats—and its impact on team performance. Unlike prior research which saw no significant home advantage shift during the COVID-19 pandemic, possibly due to generally lower attendances in women's games, our analysis reveals that higher crowd density correlates with improved performance across key indicators. Additionally, by controlling for the gender of referees, we uncover that female referees might experience distinct levels of crowd pressure, affecting their officiating. These findings not only contribute to our understanding of home advantage in sports but also suggest the influence of gender dynamics on the psychological aspects of refereeing.

**JEL Classification:** C93, D91, J16
**Keywords:** Home Advantage, Gender Differences, Spectator Influence, Referee Bias.

## 3.1 Introduction

Home advantage (HA) represents a well-documented phenomenon across professional sports, with a substantial body of research dedicated to soccer (Pollard; 1986; Clarke and Norman; 1995). This advantage is attributed to several key factors: familiarity with the home venue (Nevill and Holder; 1999), travel-related fatigue affecting the visiting team (Courneya and Carron; 1991; Pace and Carron; 1992), and the psychological and motivational impact of crowd support (Agnew and Carron; 1994; Nevill et al.; 1996). Excessive fan support is particularly noted for its dual influence; it not only boosts the home team's performance by increasing player effort and overall performance (Bray and Widmeyer; 2000; Thirer and Rampey; 1979; Schwartz and Barsky; 1977) but also exerts pressure on referees. This pressure can lead to a bias, with referees potentially making more decisions in favor of the home team (Nevill and Holder; 1999).

The interplay between these elements underscores the multifaceted nature of HA. While the positive effects on home team performance are widely recognized, the subtler impacts on referee decision-making highlight the complexity of social and psychological influences at play.

The consideration of referees' unconscious biases further illuminates the intricate dynamics of home advantage. Studies have demonstrated that such biases can significantly influence match outcomes. For instance, referees are more likely to issue red cards to teams outside their linguistic community, highlighting how social and cultural factors might sway decision-making processes. Similarly, away teams often face disadvantages when competing against home teams sharing the same linguistic background as the referee, suggesting that these biases extend beyond the immediate game environment to encompass broader social identities (Krumer and Smith; 2023).

Furthermore, the variability of home advantage linked to referees' subjective decisions is supported by empirical evidence. Boyko et al. (2007), in their examination of English Premier League (EPL) matches involving 50 referees, revealed that home advantage is not a uniform phenomenon but varies significantly among referees. This variability suggests that the subjective interpretation of rules and game situations by referees can markedly influence the extent of home advantage.

Similarly, Downward and Jones (2007) observed consistent results within the context of the FA Cup matches, indicating that these patterns are not confined to league play but extend to cup competitions as well[1].

The influence of crowd size on referee behavior provides additional insights into the dynamics of home advantage. In the English Cup, studies have shown a correlation between crowd size and the likelihood of referees sanctioning the home team with a yellow card. As crowd size increases, the propensity to penalize the home team diminishes, suggesting that referees may be swayed by the presence and enthusiasm of spectators.

Further evidence of referees' subconscious biases is provided by Garicano et al. (2005), who analyzed Spanish referees and found a tendency to adjust the duration of injury time in

[1] The Football Association Challenge Cup, commonly known as the FA Cup, is the principal national soccer competition in England and the world's oldest official soccer tournament, established in 1871.

a manner that favors the home team. Specifically, referees assign less additional time when the home team leads by a narrow margin and extend the match duration when the home team is slightly behind. This pattern suggests that referees, perhaps subconsciously, seek to please the home crowd by influencing the match outcome through the allocation of injury time.

The COVID-19 pandemic has offered a distinctive opportunity for scholars to explore the dynamics of home advantage (HA) in the absence of public spectators at sporting events. Since 2020, an extensive array of studies across soccer and other sports has emerged, affirming the pivotal role of crowd support in HA. Notably, Scoppa (2021) analyzed the five major European soccer leagues, demonstrating that crowd support significantly contributes to home advantage across various performance indicators, including points, goals, and shots. Remarkably, this advantage diminishes by nearly half in matches played behind closed doors, indicating a stark reduction in HA without the presence of fans.

Furthermore, this shift in advantage also mirrors changes in referee behavior, with decisions on fouls, yellow cards, red cards, and penalties becoming substantially more balanced in the absence of audience pressure. These findings bolster the hypothesis that social pressure exerts a considerable influence on the behavior of key participants in sports, from players to referees[2].

Additionally, the impact of playing without an audience varies depending on teams' prior experiences with crowd support. Teams accustomed to performing in front of large crowds experience a more pronounced reduction in home advantage under these conditions. However, this effect tends to wane over time, suggesting that players gradually adapt to the absence of spectators (Fischer and Haucap; 2021). This phenomenon of adaptation and diminishing HA in ghost games has been corroborated by various studies across different sports and contexts (Ferraresi and Gucciardi; 2023; Cross and Uhrig; 2023; Dilger and Vischer; 2020; Destefanis et al.; 2022; Bryson et al.; 2021).

In the realm of women's sports, evidence strongly supports the presence of HA. Notably, Krumer (2017) found that the magnitude of HA in women's judo exceeds that observed in men's judo, suggesting discipline-specific dynamics of HA. Similarly, Yu et al. (2020) reported a significant HA effect in matches among equally matched teams in the Chinese women's volleyball league, highlighting the universal applicability of HA across genders and sports. Pollard and Gómez (2014) analyze 26 women's soccer leagues from 2004 to 2010, and find that the home advantage is between 51.0% and 58.8%, with an average of 54.2%.

However, recent research by Krumer and Smith (2023) presents a contrasting perspective within the context of the *Swedish Damallsvenskan* women's soccer league. Unlike patterns observed in men's soccer, this study found no decrease in home advantage during the COVID-19 period. Intriguingly, away teams incurred more yellow cards during this period compared to before the pandemic. This deviation suggests that the influence of social pressure from crowds on referees might be less pronounced in women's soccer than in men's soccer. A potential explanation for this discrepancy could be attributed to the differing levels of spectator interest in women's soccer in Sweden, which may not exert the same degree of

[2] Referencing Akerlof (1980), Bernheim (1994), and Benabou and Tirole (2006) for theories on social pressure and its effects.

psychological pressure on referees as seen in men's games.

Further complicating the understanding of HA in women's soccer, Szabó and Kerényi (2023) reveals varied impacts of the pandemic-induced lockdown across different leagues. Their analysis indicates a statistically significant change in the dynamics of yellow cards issued during the lockdown, yet this shift did not translate to a noticeable change in match outcomes. Thus, while the imposition of lockdown conditions appeared to affect certain aspects of game officiating, it did not ultimately alter the overall magnitude of home advantage in women's soccer competitions.

Diverging from conventional findings in the literature, our study makes a primary contribution by incorporating data on referee gender and the ratio of spectators to stadium capacity. This innovative approach allows us to evaluate the impact of stadium crowding percentage on referees' decisions with an articulated consideration of gender differences. We observe that home advantage cannot always be attributed solely to the presence of fans in the stadium. This suggests that other factors, possibly inherent to the game or influenced by broader social and psychological dynamics, also play a significant role in shaping home advantage.

The organization of the remainder of this work is as follows: the forthcoming section provides a detailed description of the data utilized in our study and presents some descriptive statistics to lay the groundwork for our analysis. Section 3 delves into the results concerning team performance indicators, offering insights into how these metrics vary across different contexts of spectator presence and referee gender. Section 4 is dedicated to examining the outcomes of decisions made by referees, with a focus on identifying gender-based patterns. Finally, Section 5 concludes our study, summarizing key findings, discussing implications, and suggesting directions for future research.

## 3.2 Data and Descriptive Statistics

To examine whether there is a difference in home advantage for women in both player and referee roles, we utilize a dataset comprising matches from the 2018-19 season to the 2022-23 season. Our analysis is centered on some of the premier European leagues, which are widely regarded for their competitive parity and high level of play. Specifically, the dataset includes a total of 2,498 matches from the following leagues: Division 1 Féminine in France (FRA), Frauen-Bundesliga in Germany (GER), Women's Serie A in Italy (ITA), and the Women's Super League in England (ENG). The selection of these leagues allows for a robust examination of home advantage across varied cultural and footballing contexts within Europe, providing a comprehensive view of its dynamics in women's soccer at the elite level.

The delineation of our dataset to the specified seasons is influenced by two primary considerations. Initially, the availability of comprehensive data for women's soccer leagues has been notably limited in comparison to men's leagues. This scarcity has only recently been mitigated, thanks to a surge in interest in women's soccer, enabling access to high-quality data for in-depth analysis[3]. Additionally, the advent of the COVID-19 pandemic in

[3] Data was sourced from FBstat.com and SoccerDonna.de, which are among the leading repositories of women's soccer statistics.

2020 introduced unprecedented variables into the sporting world, including the suspension of leagues and the subsequent resumption of matches in empty stadiums. These unique circumstances offer an invaluable natural experiment to examine the influence of spectator presence on team performance and referee behavior, further justifying the selection of this particular timeframe for our study.

Table 3.1 offers a detailed breakdown of the dataset, showcasing the distribution of matches across the selected leagues and delineating the attendance scenarios—specifically, matches played with and without spectators. Of the total 2,498 games analyzed, 504 were conducted in the absence of a live audience, representing approximately 20% of the dataset. Additionally, the table introduces the 'Crowd ratio' variable, a key metric for our analysis. This variable is defined as the ratio of actual attendance to the stadium's capacity, providing a measure of stadium utilization during matches. We posit that the crowd ratio is a significant determinant of referee decisions, hypothesizing that higher crowd densities may exert greater psychological pressure on referees. Preliminary observations suggest that the crowd ratio is fairly consistent across the leagues under study, indicating a uniformity in attendance patterns relative to stadium capacities. This consistency allows for a more controlled comparison of the hypothesized effects across different league environments.

**Table 3.1:** *Data and Attendance Composition*

| Competition | Country | N. Games | N. Games Closed Doors | Mean Attendance[a] | Mean Capacity | Mean Crowd Ratio[a] | Max Attendance |
|---|---|---|---|---|---|---|---|
| Division 1 Féminine | FRA | 624 | 102 | 864.00 | 6917.00 | 0.21 | 30661.00 |
| Frauen-Bundesliga | GER | 660 | 140 | 1203.00 | 9011.00 | 0.19 | 23200.00 |
| Serie A Femminile | ITA | 621 | 132 | 523.00 | 7333.00 | 0.18 | 39000.00 |
| Women Super League | ENG | 593 | 130 | 2897.00 | 11716.00 | 0.25 | 47367.00 |
| Total | - | 2498 | 504 | 1441.00 | 8716.00 | 0.21 | 47367.00 |

Notes: [a]Both Mean Attendance and Mean Crowd ratio are computed only on the matches with Open Doors.

To assess whether the absence of fans has differential gender effects, our dataset includes the key variables of the referee's gender. Notably, female referees account for approximately 70% of the total.

Table 3.2 reveals noteworthy disparities in terms of home advantage for both variables related to the home team's performance as well as those related to the referee's decisions. Firstly, there is a statistically significant advantage in terms of points gained ($t = 3.7068$, $p < 0.001$), indicating that home teams tend to accumulate more points on average. Additionally, they demonstrate a significant edge in scoring goals ($t = 4.1431$, $p < 0.001$). Conversely, the analysis of referee decisions reveals that home teams receive fewer yellow cards ($t = -3.7671$, $p < 0.001$), suggesting a possible bias in favor of the home team by referees. However, the difference in red cards ($t = -1.302$, $p = 0.193$) is not statistically significant. Lastly, when it comes to penalty kicks ($t = 2.114$, $p = 0.035$), home teams are more likely to be awarded penalties, indicating another facet of the home advantage. The analysis of fouls ($t = -1.4074$, $p = 0.1595$) did not yield a significant difference.

Furthermore, we calculate two distinct performance metrics for each variable: the seasonal aggregate and the performance within the last four games. These metrics serve the purpose

**Table 3.2:** *Descriptive Statistics*

| Variable | Obs | Mean | SD | Min | Max |
|---|---|---|---|---|---|
| Female referees | 2498 | 0.70 | 0.46 | 0 | 1 |
| Points Home | 2498 | 1.53 | 1.39 | 0 | 3 |
| Points Away | 2498 | 1.32 | 1.38 | 0 | 3 |
| Goals Home | 2498 | 1.76 | 1.81 | 0 | 11 |
| Goals Away | 2498 | 1.53 | 1.63 | 0 | 14 |
| Red Cards Home | 2496 | 0.04 | 0.21 | 0 | 3 |
| Red Cards Away | 2496 | 0.05 | 0.22 | 0 | 2 |
| Yellow Cards Home | 2496 | 1.13 | 1.06 | 0 | 6 |
| Yellow Cards Away | 2496 | 1.24 | 1.10 | 0 | 6 |
| Fouls committed Home | 1381 | 10.09 | 4.06 | 0 | 26 |
| Fouls committed Away | 1381 | 10.27 | 3.92 | 0 | 34 |
| Penalty Kicks Awarded Home | 2496 | 0.12 | 0.33 | 0 | 2 |
| Penalty Kicks Awarded Away | 2496 | 0.10 | 0.31 | 0 | 2 |
| Points latest 4 Home | 2355 | 5.73 | 3.65 | 0 | 12 |
| Points latest 4 Away | 2357 | 5.79 | 3.62 | 0 | 12 |
| Goals latest 4 Home | 2355 | 6.59 | 4.55 | 0 | 28 |
| Goals latest 4 Away | 2357 | 6.65 | 4.57 | 0 | 26 |

Notes: missing values for Shoots, Shoots on Target, and Fouls are primarily attributable to data from the Serie A Femminile (ITA), Frauen-Bundesliga (GER) and Division 1 Féminine (FRA).

of controlling for the form status of the teams [4].

In our analysis, we adopt a dual-perspective approach, aligning with the methodologies employed by Garicano et al. (2005); Ponzo and Scoppa (2018); Scoppa (2021). Specifically, we examine each match from two viewpoints: that of the home team and that of the away team. This approach allows us to account for the 'Home' dummy variable and isolate the influence of playing at home while considering the presence or absence of spectators. Finally, we cluster standard errors at the match level to provide a more robust framework for our estimations.

## 3.3 Home Advantage and Team Performance

Our exploration into the presence of home advantage begins with an analysis focused on two key performance metrics: points gained and goals scored during the competition. These metrics are intrinsically linked, as scoring more goals directly contributes to a team's ability to secure points, thus improving its standing in the competition rankings. Despite this correlation, it is crucial to dissect the relationship between these variables to understand if distinct dynamics influence the attainment of points versus the scoring of goals.

### 3.3.1 Points

The regression analysis detailed in Table 3.3 systematically investigates the impact of home advantage on points accumulation across different match settings. This examination underscores the consistent performance superiority of home teams, suggesting that home advantage is not solely contingent upon the presence of a crowd.

[4] It's worth noting that we have calculated performance trends for each variable, although they are not included in the table.

Our regression analysis, as summarized in Table 3.3, explores the influence of home advantage on points accrued under various match conditions, affirming the consistent advantage for home teams. This advantage persists regardless of crowd presence, suggesting factors beyond mere spectator support contribute to home superiority.

The regression coefficients for 'Home' in models (I), (II), and (III) affirm a steady home advantage, with home teams earning more points on average. Specifically, teams at home secure an additional 0.175 points in scenarios with spectators (model I) and 0.232 points in matches played without spectators (model II). Intriguingly, the absolute crowd size (measured in thousands) does not significantly affect the points earned or the home team's advantage. Model (III) further indicates that the absence of spectators ('Closed Doors') does not markedly diminish home advantage, as evidenced by the small and statistically insignificant coefficient.

Adjusting for team performance and disciplinary actions in models (IV) through (VI) refines our understanding[5]. However, the relationship between crowd size and home team points gains a new dimension when considering 'Crowd Ratio'—the ratio of spectators to stadium capacity—in models (VII) and (VIII). The significant positive interaction between 'Home' and 'Crowd Ratio' (0.995 and 0.385, respectively) in these models reveals that denser crowds, rather than larger ones, are significantly associated with increased home team points. This finding suggests that smaller, more proportionately filled stadiums may offer a more potent home advantage in women's soccer leagues. It implies that the intimacy and intensity of the crowd, achievable in appropriately sized venues, can be more influential than the only number of spectators.

This effect underscores the importance of considering stadium capacity relative to attendance in analyzing home advantage, particularly in contexts where women's sports traditionally attract smaller crowds.

### 3.3.2 Goals

Given the observed influence of the 'Crowd ratio' on team point performance, our next line of inquiry delves into whether this effect extends to the goals scored by teams, as detailed in Table 3.4. This exploration is critical in understanding the broader impact of audience presence on the advantage in scoring a goal since it is the main objective of the players during the match.

The 'Home' variable demonstrates a statistically significant positive impact on goals scored across various models, highlighting a robust home advantage. For instance, teams playing at home score an additional 0.239 goals ($p<0.01$) in open-door scenarios (column I), affirming the substantial benefit of home play. However, this home scoring advantage appears diminished and statistically non-significant under closed-door conditions, with a coefficient reduction to 0.163 in column II. The variable 'Closed Doors' itself exhibits a minimal and statistically non-significant effect on goal scoring, suggesting that the intrinsic advantages of home play persist irrespective of spectator presence.

[5] Control variables for recent performance include the sum of points and goals from the last four games; disciplinary actions are captured through red and yellow cards received.

**Table 3.3:** *Home Advantage: Points*

| | *Dependent variable:* N. Points | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | (I) Open Doors | (II) Closed Doors | (III) All | (IV) Open Doors | (V) Closed Doors | (VI) All | (VII) Crowd Ratio | (VIII) Crowd Ratio |
| **Home** | **0.175**** | **0.232*** | **0.175**** | **0.225***** | **0.223**** | **0.223***** | **0.056** | **0.177***** |
| | (0.072) | (0.124) | (0.072) | (0.064) | (0.104) | (0.064) | (0.076) | (0.065) |
| Attendance (1000) | −0.013 | | −0.013 | −0.008 | | −0.007 | | |
| | (0.008) | | (0.008) | (0.008) | | (0.008) | | |
| Crowd ratio | | | | | | | −0.475*** | −0.188 |
| | | | | | | | (0.144) | (0.122) |
| Closed Doors | | | −0.028 | | | −0.005 | 0.006 | 0.001 |
| | | | (0.072) | | | (0.061) | (0.010) | (0.011) |
| Home x Attendance (1000) | 0.026 | | 0.026 | 0.013 | | 0.013 | | |
| | (0.017) | | (0.017) | (0.016) | | (0.016) | | |
| Home x Closed Doors | | | 0.057 | | | 0.007 | | |
| | | | (0.144) | | | (0.122) | | |
| **Home x Crowd ratio** | | | | | | | **0.995***** | **0.385*** |
| | | | | | | | (0.289) | (0.183) |
| Season and League FE | No | No | No | Yes | Yes | Yes | No | Yes |
| Control Variables | No | No | No | Yes | Yes | Yes | No | Yes |
| Observations | 3,278 | 982 | 4,260 | 2,970 | 954 | 3,924 | 4,260 | 3,924 |
| Adjusted $R^2$ | 0.006 | 0.006 | 0.006 | 0.309 | 0.322 | 0.313 | 0.011 | 0.314 |

*Note:* OLS estimates. Robust standard errors are reported in parentheses and adjusted for clustering at the match level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Interestingly, the initial analysis reveals a significant effect of attendance on home team goals, with the 'Attendance' variable initially showing a positive association (column I) that becomes statistically non-significant upon the addition of control variables (column IV). This shift indicates that while overall attendance might initially seem to impact home team scoring, its effect diminishes when controlling for team form and disciplinary actions. Moreover, the interaction term 'Home x Closed Doors' suggests that the combined effect of playing at home and without spectators does not significantly alter the home advantage in terms of goals scored.

Lastly, columns VII and VIII provide a more precise measure of crowd influence by considering audience size relative to stadium capacity. Here, we can observe a significant increase in home team goals as stadium occupancy approaches full capacity. Column VII shows a notable coefficient of 1.33, and this slightly decreases to 0.559 in column VIII after controlling for league and season effects, as well as past team performance.

These findings suggest that smaller, more proportionally filled venues may offer strategic advantages in women's soccer by intensifying the home advantage effect. In other words, a dense, engaged audience can significantly contribute to the home advantage, possibly by creating a more intimidating atmosphere for visiting teams or bolstering home team morale.

## 3.4 Home Advantage and Referee's Gender

In the preceding section, we explored the home advantage's impact on points and goals in women's football, uncovering interesting findings. The significant positive effects found in

**Table 3.4:** *Home Advantage: Goals*

| | *Dependent variable:* N. Goals | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | (I) Open Doors | (II) Closed Doors | (III) All | (IV) Open Doors | (V) Closed Doors | (VI) All | (VII) Crowd Ratio | (VIII) Crowd Ratio |
| **Home** | **0.239**$^{***}$ | **0.175** | **0.239**$^{***}$ | **0.293**$^{***}$ | **0.162** | **0.294**$^{***}$ | **0.045** | **0.198**$^{***}$ |
| | (0.074) | (0.129) | (0.074) | (0.062) | (0.100) | (0.062) | (0.079) | (0.065) |
| **Attendance (1000)** | $-$**0.022**$^{***}$ | | $-$**0.022**$^{***}$ | $-$**0.017**$^{*}$ | | $-$**0.017**$^{*}$ | | |
| | (0.007) | | (0.007) | (0.009) | | (0.009) | | |
| Crowd ratio | | | | | | | $-0.550^{***}$ | $-0.278^{*}$ |
| | | | | | | | (0.171) | (0.162) |
| Closed Doors | | | $-0.017$ | | | 0.004 | $-0.016$ | $-0.048$ |
| | | | (0.088) | | | (0.079) | (0.059) | (0.060) |
| Home x Attendance (1000) | $0.032^{**}$ | | $0.032^{**}$ | 0.022 | | 0.022 | | |
| | (0.015) | | (0.015) | (0.014) | | (0.014) | | |
| Home x Closed Doors | | | $-0.064$ | | | $-0.124$ | | |
| | | | (0.149) | | | (0.118) | | |
| **Home x Crowd ratio** | | | | | | | **1.330**$^{***}$ | **0.559**$^{**}$ |
| | | | | | | | (0.327) | (0.254) |
| Season and League FE | No | No | No | Yes | Yes | Tes | No | Yes |
| Control Variables | No | No | No | Yes | Yes | Yes | No | Yes |
| Observations | 3,988 | 1,008 | 4,996 | 3,640 | 978 | 4,618 | 4,260 | 3,924 |
| Adjusted $R^2$ | 0.005 | 0.001 | 0.004 | 0.249 | 0.272 | 0.255 | 0.011 | 0.271 |

*Note:* OLS estimates. Robust standard errors are reported in parentheses and adjusted for clustering at the match level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

our models on teams' performances, especially those accounting for crowd ratio interactions. The results suggest that home advantage is less about the mere presence of fans and more about their proportion relative to stadium capacity. Now we focus on analyzing if there are effects on the behavior of the referees and the presence of a gender difference.

### 3.4.1 Penalty Kicks

Table **??** offers an analysis of factors influencing the awarding of penalty kicks in women's soccer through linear regression. Contrary to initial assumptions, a distinct home advantage in penalty kick awarding is not evident. Notably, a clear home advantage in penalty awarding emerges only under certain conditions, with significant findings in models IV and VII suggesting that home teams are marginally more likely to receive penalty kicks.

Interestingly, an examination of the findings in column VI reveals a notable gender disparity in the issuance of penalty kicks. Specifically, during matches played in the absence of a crowd, the data indicates an overall increase in the number of penalty kicks awarded. However, it is observed that female referees tend to award fewer penalty kicks, with a coefficient of -0.075 and a significance level of $p < 0.10$. This pattern suggests the possibility of a distinct officiating approach based on referee gender, presenting an intriguing avenue for future research to explore further.

The introduction of control variables, including season and league specifics, along with teams' recent performance metrics, does not markedly alter the Adjusted R-squared values. This stability suggests that the observed effects of home advantage and referee gender are robust across different league contexts and independent of team form. Attempts to include referee-fixed effects further confirmed the consistency of these findings.

The examination of crowd ratio impacts reveals a significant negative correlation between crowd density and the likelihood of penalty awards (-0.157 and -0.164), suggesting that denser crowds may deter the awarding of penalties. This trend persists irrespective of home or away team status, adding a layer of complexity to the understanding of home advantage and referee behavior.

**Table 3.5:** *Referee's gender and Home Advantage: Penalty Kicks*

| | *Dependent variable:* **N. Penalty Kicks** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | (I) Open Doors | (II) Closed Doors | (III) All | (IV) Open Doors | (V) Closed Doors | (VI) All | (VII) Crowd Ratio | (VIII) Crowd Ratio |
| **Home** | **0.023** | **0.029** | **0.023** | **0.032**$^{*}$ | **−0.002** | **0.032**$^{*}$ | **0.018** | **0.015** |
| | (0.017) | (0.053) | (0.017) | (0.018) | (0.051) | (0.018) | (0.029) | (0.030) |
| Closed Doors | | | 0.085$^{**}$ | | | **0.095**$^{**}$ | | |
| | | | (0.037) | | | (0.038) | | |
| Crowd Ratio | | | | | | | −0.157$^{***}$ | −0.164$^{***}$ |
| | | | | | | | (0.053) | (0.059) |
| **Female** | **−0.009** | **−0.071**$^{*}$ | **−0.009** | **0.011** | **−0.011** | **0.021** | **−0.028** | **−0.006** |
| | (0.015) | (0.038) | (0.015) | (0.022) | (0.062) | (0.021) | (0.022) | (0.029) |
| Home x Closed Doors | | | 0.006 | | | −0.027 | | |
| | | | (0.056) | | | (0.054) | | |
| Home x Crowd Ratio | | | | | | | 0.014 | 0.047 |
| | | | | | | | (0.096) | (0.102) |
| Home x Female | 0.0004 | −0.037 | 0.0004 | −0.014 | −0.010 | −0.014 | −0.031 | −0.026 |
| | (0.021) | (0.058) | (0.021) | (0.021) | (0.056) | (0.021) | (0.032) | (0.033) |
| **Closed Doors x Female** | | | −0.062 | | | **−0.075**$^{*}$ | | |
| | | | (0.041) | | | (0.042) | | |
| Home x Closed Doors x Female | | | −0.038 | | | −0.0004 | | |
| | | | (0.062) | | | (0.060) | | |
| Crowd ratio x Female | | | | | | | 0.055 | 0.075 |
| | | | | | | | (0.061) | (0.066) |
| Home x Crowd ratio x Female | | | | | | | 0.154 | 0.085 |
| | | | | | | | (0.109) | (0.115) |
| Season and League FE | No | No | No | Yes | Yes | Yes | No | Yes |
| Control Variables | No | No | No | Yes | Yes | Yes | No | Yes |
| Observations | 3,988 | 1,004 | 4,992 | 3,640 | 978 | 4,618 | 4,256 | 3,924 |
| Adjusted $R^2$ | 0.001 | 0.010 | 0.004 | 0.022 | 0.049 | 0.029 | 0.004 | 0.033 |

*Note:* OLS estimates. Robust standard errors are reported in parentheses and adjusted for clustering at the match level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

### 3.4.2 Red Cards

Table 3.6 delves into the dynamics of red card issuance in women's soccer, scrutinizing the roles of crowd presence, home advantage, and the gender of referees.

Initially, the analysis reveals no substantial home advantage in matches with spectators present, as indicated by a marginal increase in red cards awarded to home teams. This suggests that the crowd's presence may primarily amplify the intensity and aggression within the game rather than exerting a direct influence on referees' decisions. However, an intriguing pattern emerges in closed-door matches, detailed in Column VI, where home teams are notably less likely to be issued red cards (-0.063), pointing towards the nuanced impact of spectator absence on officiating behaviors.

A particularly interesting finding is the distinct behavior exhibited by female referees, who statistically show a greater propensity to issue red cards to home teams (0.070), a trend that remains consistent across various controls for seasons, leagues, and team performance. This observation hints at a gender-specific approach to officiating, especially in environments devoid of spectator influence, suggesting nuanced differences in how referees of different genders perceive and respond to home advantage dynamics.

As previously, in this case too, we attempted to quantify the fans' impact on the game using the 'Attendance' variable. However, similar to earlier findings, this variable failed to yield any significant results, underscoring the challenge of capturing the effect of fans on game outcomes through attendance figures.

Interestingly, also the crowd density does not significantly impact the number of red cards issued (Columns VII and VIII). This observation suggests that factors beyond mere crowd density—such as the specific context of the game or referee discretion—play a more important role in determining disciplinary outcomes.

**Table 3.6:** *Referee's gender and Home Advantage: Red Cards*

| | *Dependent variable:* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **N. Red Cards** | | | | | | | |
| | (I) Open Doors | (II) Closed Doors | (III) All | (IV) Open Doors | (V) Closed Doors | (VI) All | (VII) Crowd Ratio | (VIII) Crowd Ratio |
| **Home** | **0.008** | **−0.051**** | **0.008** | **0.009** | **−0.054**** | **0.009** | **−0.016** | **−0.020** |
| | (0.014) | (0.026) | (0.014) | (0.015) | (0.027) | (0.015) | (0.018) | (0.018) |
| Closed Doors | | | 0.017 | | | 0.020 | | |
| | | | (0.024) | | | (0.025) | | |
| Crowd Ratio | | | | | | | 0.096 | 0.112 |
| | | | | | | | (0.080) | (0.083) |
| Female | −0.010 | −0.037 | −0.010 | 0.005 | 0.014 | 0.012 | −0.013 | 0.011 |
| | (0.011) | (0.024) | (0.011) | (0.015) | (0.043) | (0.015) | (0.015) | (0.020) |
| **Home x Closed Doors** | | | **−0.059**** | | | **−0.063**** | | |
| | | | (0.029) | | | (0.031) | | |
| Home x Crowd Ratio | | | | | | | −0.048 | −0.032 |
| | | | | | | | (0.100) | (0.099) |
| Home x Female | −0.020 | 0.048* | −0.020 | −0.019 | 0.051* | −0.019 | 0.013 | 0.019 |
| | (0.016) | (0.029) | (0.016) | (0.017) | (0.030) | (0.017) | (0.020) | (0.020) |
| Closed Doors x Female | | | −0.027 | | | −0.028 | | |
| | | | (0.027) | | | (0.028) | | |
| **Home x Closed Doors x Female** | | | **0.068**** | | | **0.070**** | | |
| | | | (0.033) | | | (0.035) | | |
| Crowd ratio x Female | | | | | | | −0.075 | −0.084 |
| | | | | | | | (0.084) | (0.086) |
| Home x Crowd ratio x Female | | | | | | | 0.027 | 0.011 |
| | | | | | | | (0.104) | (0.104) |
| Season and League FE | No | No | No | Yes | Yes | Yes | No | Yes |
| Control Variables | No | No | No | Yes | Yes | Yes | No | Yes |
| Observations | 3,988 | 1,004 | 4,992 | 3,640 | 978 | 4,618 | 4,256 | 3,924 |
| Adjusted $R^2$ | 0.002 | 0.003 | 0.002 | 0.002 | 0.012 | 0.004 | 0.002 | 0.005 |

*Note:* OLS estimates. Robust standard errors are reported in parentheses and adjusted for clustering at the match level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

### 3.4.3 Yellow Cards

Lastly, Table 3.7 offers an extensive analysis of the dynamics between referee gender, crowd presence, and the frequency of yellow cards awarded in women's soccer. Contrary to the pattern observed with red cards, home teams under open-door conditions see a statistically significant decrease in yellow cards received (-0.142), indicating a clear home advantage in this disciplinary aspect. However, this advantage reverses in closed-door matches, although not significantly so (Column II), suggesting that the absence of spectators might subtly alter referees' tendencies toward caution issuance.

The analysis does not reveal a significant overall trend for female referees issuing more or fewer yellow cards to home teams in either open or closed matches. The interaction between referee gender and closed doors (Columns V and VI) indicates no significant change in yellow card issuance, suggesting that female referees' decision-making regarding yellow cards may not be distinctly influenced by the presence of spectators. Introducing control variables, including season, league specifics, and team performance (Columns IV to VI), slightly modify the observed patterns, indicating that broader match contexts and team dynamics play a role in yellow card issuance. Again, the crowd ratio analysis (Columns VII and VIII) shows no significant direct impact on yellow cards, pointing to the complexity of factors influencing referees' decisions beyond simple crowd density.

Notably, the interaction between home advantage and crowd ratio significantly influences yellow card issuance in the expected direction, with higher crowd densities correlated with fewer yellow cards for home teams (-0.661 and -0.506), reinforcing the notion of home advantage in the context of referee decisions under varying crowd densities.

**Table 3.7:** *Referee's gender and Home Advantage: Yellow Cards*

| | *Dependent variable:* **N. Yellows** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | (I) Open Doors | (II) Closed Doors | (III) All | (IV) Open Doors | (V) Closed Doors | (VI) All | (VII) Crowd Ratio | (VIII) Crowd Ratio |
| **Home** | $-$**0.142**$^{***}$ | **0.029** | $-$**0.142**$^{***}$ | $-$**0.125**$^{**}$ | **0.034** | $-$**0.125**$^{**}$ | $-$**0.049** | $-$**0.068** |
| | (0.055) | (0.112) | (0.055) | (0.058) | (0.116) | (0.059) | (0.076) | (0.081) |
| Closed Doors | | | $-0.093$ | | | $-0.095$ | | |
| | | | (0.099) | | | (0.103) | | |
| Crowd Ratio | | | | | | | 0.041 | 0.084 |
| | | | | | | | (0.243) | (0.230) |
| Female | 0.045 | $-0.019$ | 0.045 | 0.026 | 0.060 | 0.047 | 0.091 | 0.020 |
| | (0.052) | (0.106) | (0.052) | (0.085) | (0.175) | (0.080) | (0.069) | (0.096) |
| Home x Closed Doors | | | 0.171 | | | 0.164 | | |
| | | | (0.125) | | | (0.129) | | |
| Home x Crowd Ratio | | | | | | | $-0.661^{**}$ | $-0.506^{*}$ |
| | | | | | | | (0.306) | (0.289) |
| Home x Female | $-0.005$ | 0.018 | $-0.005$ | $-0.018$ | 0.008 | $-0.017$ | 0.015 | 0.023 |
| | (0.067) | (0.135) | (0.067) | (0.071) | (0.137) | (0.071) | (0.089) | (0.093) |
| Closed Doors x Female | | | $-0.063$ | | | $-0.019$ | | |
| | | | (0.118) | | | (0.121) | | |
| Home x Closed Doors x Female | | | 0.022 | | | 0.028 | | |
| | | | (0.151) | | | (0.154) | | |
| Crowd ratio x Female | | | | | | | 0.128 | 0.052 |
| | | | | | | | (0.277) | (0.266) |
| Home x Crowd ratio x Female | | | | | | | 0.263 | 0.224 |
| | | | | | | | (0.354) | (0.338) |
| Season and League FE | No | No | No | Yes | Yes | Yes | No | Yes |
| Control Variables | No | No | No | Yes | Yes | Yes | No | Yes |
| Observations | 3,988 | 1,008 | 4,996 | 3,640 | 982 | 4,622 | 4,260 | 3,928 |

*Note:* OLS estimates. Robust standard errors are reported in parentheses and adjusted for clustering at the match level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

These outcomes suggest that the home advantage, in terms of decisions taken by the referees, is a multifaceted phenomenon. It is influenced by a combination of environmental and psychological factors rather than solely by the physical presence of spectators or the gender of the referee.

## 3.5 Conclusion

This study has embarked on a detailed exploration of home advantage in women's soccer, leveraging the unique circumstances brought about by COVID-19 to dissect the interplay between crowd presence, referee gender, and team performance. Despite the burgeoning interest in home advantage across sports disciplines, women's soccer remains underexplored, with only a handful of studies, such as those by Krumer and Smith (2023) and Szabó and Kerényi (2023), beginning to shed light on this phenomenon in recent times.

The core findings of our study enrich the understanding of home advantage in women's soccer leagues in several key ways. Our analysis decisively shows that the number of points garnered by home teams is not directly influenced by the sheer size of attendance. Instead, it is the crowd ratio—the proportion of fans relative to stadium capacity—that significantly impacts the points earned by home teams. This highlights the strategic importance of venue size and crowd density, suggesting that smaller, optimally filled stadiums can enhance home advantage by creating a more engaged and potentially intimidating atmosphere for visiting teams.

In contrast to the points received, the size of attendance does influence the home advantage in scoring a goal, indicating that the presence and magnitude of the crowd bear a significant impact on team performance beyond just winning points. This finding increases our understanding of how different aspects of audience presence can variably affect home team outcomes.

Regarding referee decisions, our study uncovers gender-specific patterns in officiating that vary with match conditions. In games played without spectators, female referees are found to award fewer penalty kicks and red cards. This suggests a potential gender-related difference in response to the absence of crowd pressure, with female referees possibly exhibiting a more cautious approach in high-stakes decisions. Additionally, both male and female referees tend to issue fewer yellow cards to home teams in closed-door matches. This uniformity across genders implies a broader officiating trend that transcends individual referee gender, reflecting perhaps a subconscious bias or a shift in game dynamics when external pressures are minimized.

In synthesizing these observations, our research not only contributes to the expanding literature on home advantage but also introduces new dimensions of analysis—such as the impact of crowd ratio and the role of referee gender in decision-making under varied conditions. These insights encourage a reevaluation of strategies regarding venue selection and highlight the need for further exploration into the behavioral underpinnings of referee decisions in sports.

Future inquiries may extend beyond the scope of women's soccer to ascertain if these patterns hold across other sports or different competitive environments. Such research could offer broader implications for understanding the interplay between environmental factors, gender dynamics, and performance in high-pressure situations.

## References

Agnew, G. and Carron, A. (1994). Crowd effects and the home advantage, *International Journal of Sport Psychology* .

Akerlof, G. (1980). A theory of social custom, of which unemployment may be one consequence, *Quarterly Journal of Economics* **94**: 749–775.

Benabou, R. and Tirole, J. (2006). Incentives and prosocial behavior, *American Economic Review* **96**: 1652–78.

Bernheim, D. (1994). A theory of conformity, *Journal of Political Economy* **102**: 841–87.

Boyko, R. H., Boyko, A. R. and Boyko, M. G. (2007). Referee bias contributes to home advantage in english premiership football, **25**(11): 1185–1194.

Bray, S. and Widmeyer, W. (2000). Athletes' perceptions of the home advantage: An investigation of perceived causal factors, *Journal of Sport Behavior* **23**(1): 1–1.

Bryson, A., Dolton, P., Reade, J., Schreyer, D. and Singleton, C. (2021). Causal effects of an absent crowd on performances and refereeing decisions during covid-19, *Economics Letters* **198**: 109664.

Clarke, S. and Norman, J. (1995). Home ground advantage of individual clubs in english soccer, *Journal of the Royal Statistical Society: Series D (The Statistician)* **44**(4): 509–521.

Courneya, K. and Carron, A. (1991). Effects of travel and length of home stand/road trip on tie home advantage, *Journal of Sport and Exercise Psychology* **13**(1): 42–49.

Cross, J. and Uhrig, R. (2023). Do fans impact sports outcomes? a covid-19 natural experiment, *Journal of Sports Economics* **24**(1): 3–27.

Destefanis, S., Addesa, F. and Rossi, G. (2022). The impact of covid-19 on home advantage: a conditional order-m analysis of football clubs' efficiency in the top-5 european leagues, *Applied Economics* **54**(58): 6639–6655.

Dilger, A. and Vischer, L. (2020). No home bias in ghost games.

Downward, P. and Jones, M. (2007). Effects of crowd size on referee decisions: Analysis of the fa cup, *Journal of sports sciences* **25**(14): 1541–1545.

Ferraresi, M. and Gucciardi, G. (2023). Team performance and the perception of being observed: Experimental evidence from top-level professional football, *German Economic Review* **24**(1): 1–31.

Fischer, K. and Haucap, J. (2021). Does crowd support drive the home advantage in professional football? evidence from german ghost games during the covid-19 pandemic, *Journal of Sports Economics* **22**(8): 982–1008.

Garicano, L., Palacios-Huerta, I. and Prendergast, C. (2005). Favoritism under social pressure, **87**(2): 208–216.

Krumer, A. (2017). On winning probabilities, weight categories, and home advantage in professional judo, *Journal of Sports Economics* **18**(1): 77–96.

Krumer, A. and Smith, V. (2023). The effect of covid-19 on home advantage in women's soccer: Evidence from swedish damallsvenskan, *American Behavioral Scientist* **67**(10): 1168–1178.

Nevill, A. and Holder, R. (1999). Home advantage in sport: An overview of studies on the advantage of playing at home, *Sports Medicine* **28**: 221–236.

Nevill, A., Newell, S. and Gale, S. (1996). Factors associated with home advantage in english and scottish soccer matches, *Journal of Sports Sciences* **14**(2): 181–186.

Pace, A. and Carron, A. (1992). Travel and the home advantage, *Canadian Journal of Sport Sciences= Journal Canadien des Sciences du Sport* **17**(1): 60–64.

Pollard, R. (1986). Home advantage in soccer: A retrospective analysis, *Journal of Sports Sciences* **4**(3): 237–248.

Pollard, R. and Gómez, M. (2014). Comparison of home advantage in men's and women's football leagues in europe, *European Journal of Sport Science* **14**: S77–S83.

Ponzo, M. and Scoppa, V. (2018). Does the home advantage depend on crowd support? evidence from same-stadium derbies, *Journal of Sports Economics* **19**(4): 562–582.

Schwartz, B. and Barsky, S. (1977). The home advantage, *Social forces* **55**(3): 641–661.

Scoppa, V. (2021). Social pressure in the stadiums: Do agents change behavior without crowd support?, *Journal of economic psychology* **82**: 102344.

Szabó, D. and Kerényi, P. (2023). The causal impacts of empty stadiums on women's sports activities: Evidence from european football leagues, *Psychology of Sport and Exercise* **66**: 102385.

Thirer, J. and Rampey, M. (1979). Effects of abusive spectators' behavior on performance of home and visiting intercollegiate basketball teams, *Perceptual and Motor Skills* **48**($3_suppl)$ : 1047 − −1053.

Yu, Y., García-De-Alcaraz, A., Cui, K. and Liu, T. (2020). Interactive effects of home advantage and quality of opponent in chinese women's volleyball association league, *International Journal of Performance Analysis in Sport* **20**(1): 107–117.