Reimplementation Proposal of "Testing Mixed-Strategy Equilibria When Players Are Heterogeneous: The Case of Penalty Kicks in Soccer"

Group 40

Manuel Navalho - 113356
Rodrigo Arede - 102606
Tomás Alves - 113202

1 Introduction

1.1 Motivation

Football (soccer) is the world's most popular sport, played by over 200 million people in nearly 200 countries. Its global appeal was evident in the 2022 FIFA World Cup final between Argentina and France, which drew an audience of about 1.5 billion viewers - about 20% of the world's population. One of the most dramatic and decisive moments during a football game is the penalty kick. It occurs when a foul is committed inside the penalty area, granting the attacking team a direct shot at goal from a spot 11 meters away, with only the goalkeeper to beat. This is not only high-stakes but also psychologically intense and tactically nuanced. To the casual viewer, it may appear as a straightforward duel between the kicker and a goalkeeper. However, this apparent simplicity conceals a complex strategic interaction grounded in game theory. Each participant must anticipate the other's actions while masking their own intentions, creating an ideal setting for studying decision-making under uncertainty in high-stakes, zero-sum environments. This project proposal aims to reproduce and critically evaluate the study of Chiappori, Levitt, and Groseclose (2002), which applied game-theoretic principles—specifically the concept of mixed-strategy Nash equilibrium to real-world penalty kick data. A key contribution of their work is the consideration of player heterogeneity, a factor often abstracted away in theoretical models but crucial in empirical analysis.

1.2 Related work

The empirical study of strategic interactions in sports has gained momentum in recent decades. Notable early work includes Walker and Wooders (2001) on tennis serves, which until 2002 was the only paper with a similar spirit to the one we are replicating. Chiappori et al.'s paper stands out as a landmark study because it focuses on penalty kicks, where the discrete choices, repetitive nature, and availability of data allow a clear empirical test of the mixed-strategy equilibrium concept in a real-life environment. Subsequent research has expanded on this foundation, examining strategic interactions in baseball (Pitcher-batter duels), basketball (free throw shooting), and other sports. More recent work has also incorporated behavioral and psychological dimensions, such as the effects of pressure, loss aversion, and how that impacts player's decisions.

1.3 Problem definition and relevance

The central question addressed by this project is whether professional footballers' behavior during penalty kicks aligns with the predictions of mixed-strategy Nash equilibrium when heterogeneity is taken into account. Specifically, do players choose their strategies so that opponents are unable to exploit predictable patterns, even when players differ in abilities or tendencies? This question is not only central to sports analytics but also provides a rare opportunity to empirically validate a foundational concept of non-cooperative game theory. It examines whether real-world agents, under pressure and in competitive environments—are capable of sophisticated strategic reasoning. The project tests the robustness of equilibrium predictions beyond laboratory settings, exploring how individual characteristics such as preferred foot, historical direction choices, and psychological factors (whether the penalty occurs during a lead or deficit, at home or away, or under late-game pressure) influence strategic behavior.

1.4 Objectives

The primary objective of this project is to replicate the theoretical and empirical analysis presented by Chiappori, Groseclose, and Levitt (2002) on testing mixed-strategy equilibrium in football penalty kicks. Specifically, the main goals are:

- **Replication of Empirical Findings:** Reproduce the key statistical tests from the original study to check if players' actions match what game theory predicts in a mixed-strategy Nash equilibrium.
- **Validation of Theoretical Predictions:** Confirm whether key theoretical implications hold in the data, including:
  - The independence of kicker and goalie strategies.
  - Scoring probabilities are equal across directions.
  - Predicted distributions of actions (kickers choosing center more often than goalies stay center).
- **Evaluation of Heterogeneity:** Examine how differences among players—such as skill levels, preferred strategies, or behavioral tendencies—affect whether equilibrium predictions hold in practice
- **Testing Model Assumptions:** Assess the plausibility of critical model assumptions, including:
  - The simultaneous-move nature of the game.
  - Goalkeeper homogeneity (Assumption IG).
  - Side and Center (SC), Natural Side (NS), and Kicker's Side (KS) conditions on scoring probabilities.

- **Demonstration of Reproducibility:** Establish the replicability of one of the most prominent real-world tests of mixed strategies outside the lab, supporting broader applications of game-theoretic models.

2 Approach

2.1 Environment

The penalty kick is modeled as a two-player, zero-sum, simultaneous-move game with the following components:

Kicker: Chooses to shoot Left (L), Middle (C), or Right (R) (relative to their natural foot).

Goalkeeper (Goalie): Chooses to dive L, C, or R. Action Space: Discrete and finite (3×3 matrix).

Payoffs: Kicker aims to maximize scoring probability. Goalie aims to minimize it (zero-sum).

Information Structure: Players act simultaneously (no observation of the opponent's current move). Past strategies are common knowledge (players track historical tendencies). Why this design?

Matches the original paper's game-theoretic framework. The small strategy space allows tractable equilibrium analysis while capturing real-world complexity. Simultaneity is empirically validated in the paper (no serial correlation in actions).

Table 1: Kicker's and Goalkeeper's Frequency and Scoring Rate

Concept	# Obs.	Kicker's Frequency			Goalkeeper's Frequency			Scoring Rate
		NS	C	OS	NS	C	OS	
All penalty kicks	459	.4466	.1721	.3813	.5643	.0240	.4118	.7495
Italy	242	.4091	.1653	.4256	.5413	.0207	.4380	.7353
France	217	.4885	.1797	.3318	.5899	.0276	.3825	.7650
Home-team kickers	304	.4276	.1842	.3882	.5493	.0132	.4375	.7500
Away-team kickers	155	.4839	.1484	.3677	.5935	.0452	.3613	.7484
Right-footed kickers	384	.4375	.1667	.3958	.5703	.0234	.4063	.7656
Left-footed kickers	75	.4933	.2000	.3067	.5333	.0267	.4400	.6667
RR cases	208	.5433	.0000	.4567	.6154	.0000	.3846	.7644
GR cases	251	.3665	.3147	.3187	.5219	.0438	.4343	.7371
Right footed / RR	171	.5205	.0000	.4795	.6140	.0000	.3860	.7661
Right footed / GR	213	.3709	.3005	.3286	.5352	.0423	.4225	.7653
Left footed / RR	37	.6486	.0000	.3514	.6216	.0000	.3784	.7568
Left footed / GR	38	.3421	.3947	.2632	.4474	.0526	.5000	.5789

NOTE: NS = natural side; C = center; OS = opposite side; RR = restricted randomization; GR = general randomization.
SOURCE: Table adapted from [2], elaborated based on data from [1].

2.2 Multi-Agent System

The system consists of:

(1) **Heterogeneous Kickers:** Differ in skill (footedness, accuracy), leading to match-specific payoff matrices.
(2) **Homogeneous Goalies:** Under Assumption IG, goalies are treated as identical (supported by empirical tests).

The agents interact through a shared payoff matrix based on empirical scoring probabilities. Each penalty kick is an independent game instance. Players randomize strategies to make opponents indifferent (Nash equilibrium). Why this setup?

Heterogeneity: Reflects real-world variation in player abilities.

Homogeneity (for goalies): Simplifies aggregation; validated via fixed-effects tests in the original paper.

2.3 System Architecture

Our replication is structured as follows:

- **Data Layer** Use the original dataset of 459 penalty kicks from French/Italian leagues (1997-2000). This dataset includes:
  - Actions (L/C/R for kicker/goalie)
  - Outcome (goal scored: binary)
  - Game Context (time of the game, score, home/away) Pre-processing includes normalizing direction choices (reversing for left-footed players), removing invalid records, and encoding categorical features.
- **Model Layer**
  - **Game-Theoretical Model:** We represent the interaction as a 3x3 matrix game as shown in Table 2, where each cell corresponds to the empirical scoring probability given the chosen actions of the kicker and goalkeeper. The game is modeled as a zero-sum simultaneous-move game.

3 Empirical Evaluation

We test the validation of the mixed-strategy equilibrium predictions through:

3.1 Metrics explanation

Agent Strategy Estimation: We estimate each agent's mixed strategy using observed action frequencies. Kickers' and goalkeepers' empirical distributions over {L, C, R} serve as proxies for equilibrium behavior. Player heterogeneity is captured by allowing strategy variation across individuals. To account for heterogeneity, each kicker's and goalkeeper's strategy is computed separately, reflecting variation in preferences, skill, and style. This allows us to test both player-specific equilibrium properties and population-wide predictions that are robust to aggregation.

By using observed frequencies instead of structural estimation, we maintain alignment with the methodology of the original study and focus on behavioral consistency with mixed-strategy Nash equilibrium rather than exact parameter recovery.

Absence of Systematic Predictability: In equilibrium, players should randomize in such a way that their actions are not predictable. This implies that: There should be no exploitable patterns in kick direction based on historical play.

Past actions (such as the direction of the previous kick) should not predict current decisions.

We validate this by checking for serial correlation and using regressions to test whether previous outcomes or directions significantly predict future choices. A lack of predictability supports the equilibrium hypothesis.

Exploitable Patterns Test: We simulate counterfactual scenarios where kickers or goalkeepers deviate from their observed behavior to a "best response" strategy (always choosing the statistically better-performing side). If these alternative strategies produce higher payoffs than the observed strategy, it would suggest suboptimal, non-equilibrium behavior. Conversely, small or zero gains from such deviations support the equilibrium prediction.

Table 2: Generalized Payoff Matrix

		G_i		
		L	C	R
K_i	L	P_L	π_L	π_L
	C	μ	0	μ
	R	π_R	π_R	P_R

NOTE: P = Probability of scoring when the kicker and goalie choose the same side; μ = Probability of scoring when the kicker shoots for the center and the goalie jumps to one side; π = Probability of scoring when the goalie chooses the wrong side.

. **Metrics and Methods:**

- Linear Probability Model (LPM) Regression Coefficients
- Action Frequency Analysis
- Serial Correlation Test
- F-Test for Goalie Homogeneity

References

[1] Pierre-André Chiappori, Steven Levitt, and Timothy Groseclose. 2002. Testing Mixed-Strategy Equilibria When Players Are Heterogeneous: The Case of Penalty Kicks in Soccer. *American Economic Review* 92, 4 (2002), 1138–1151. doi:10.1257/00028280260344640

[2] Germán Coloma. 2007. Penalty Kicks in Soccer: An Alternative Methodology for Testing Mixed Strategy Equilibria. *Journal of Sports Economics* (2007).