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

Manuel Navalho - 113356
manuel.navalho@tecnico.ulisboa.pt
Instituto Superior Técnico
Lisboa, Portugal

Rodrigo Arêde - 102606
rodrigo.arede@tecnico.ulisboa.pt
Instituto Superior Técnico
Lisboa, Portugal

Tomás Alves - 113202
tomasmalves@tecnico.ulisboa.pt
Instituto Superior Técnico
Lisboa, Portugal

Keywords

Mixed-Strategy, Nash Equilibrium, Simultaneity, Empirical tests

1 Introduction

The concept of mixed-strategy Nash equilibrium is a cornerstone of game theory, yet its empirical validation in real-world settings remains a subject of debate. While laboratory experiments have yielded mixed results, natural environments like competitive sports offer a unique opportunity to test these theoretical predictions under high-stakes conditions. This report focuses on reproducing and critically evaluating the seminal study by Chiappori, Levitt, and Groseclose (2002) [1] , which investigates whether professional football players follow mixed-strategy equilibrium during penalty kicks, a setting that closely approximates a zero-sum, simultaneous-move game. Our primary contribution lies in reproducing the assumptions and empirical tests in the original study. First, we examine whether kickers and goalies make their decision simultaneously, a critical assumption for modeling penalty kicks as a zero-sum, simultaneous-move game. Next, we test predictions of the model that are robust to aggregation, ensuring that strategic patterns persist even when individual-level data are pooled across players. Finally, we investigate whether the goalies can be treated as behaviorally identical, a simplifying assumption with important implications for the generalizability and interpretation of the equilibrium outcomes.

2 Dataset Collection and Description

2.1 Data Sources and Compilation

Regarding the dataset, we encountered several challenges. The paper we based our work on used data from the French and Italian leagues (1997–2000), but did not publish the dataset. Additionally, the original study only provided aggregated shot frequency data, lacking detailed identifiers for players and goalies. To address these limitations, after careful consideration, we decided to compile our own dataset, consisting of 482 penalty kicks. These kicks cover every penalty taken in the English Premier League (2023–24), Portuguese League (2024–25), and Italian League (2020–21, 2022–23). The dataset was assembled by manually reviewing video highlights from official YouTube channels. For each penalty, we collected detailed information, including the identities of the kicker and goalie, their respective actions (Left, Center, or Right), the outcome of the penalty, the kicker's dominant foot, and contextual game details such as home or away status, the minute of the penalty, and the scoreline at the time of the kick. In total, the dataset includes 180 different kickers and 110 different goalies. To make our analysis comparable to that of the original paper, we focused on two subsets: (i) the 42 kickers with at least four penalty attempts (representing 58% of all observations), and (ii) the 12 kickers with at least eight attempts (accounting for 25% of the observations). This allowed us to maintain similar sample sizes to those used in the original study.

2.2 Summary Statistics

Table 1 presents an overview of the key characteristics of our compiled penalty kick dataset.

Table 1: Summary Statistics of Penalty Kick Dataset

Statistic	Count	Percentage
Total penalties	482	100.0%
Unique kickers	180	–
Unique goalies	110	–
Right-footed kickers	323	67.0%
Left-footed kickers	159	33.0%
Goals scored	401	83.2%
Missed/saved	81	16.8%
Home penalties	251	52.1%
Away penalties	231	47.9%
Game State at Kick		
Leading	131	27.2%
Drawing	223	46.3%
Trailing	128	26.6%
Minute of Match		
0–14 minutes	38	7.9%
15–29 minutes	76	15.8%
30–44 minutes	68	14.1%
45–59 minutes	93	19.3%
60–74 minutes	84	17.4%
75–90 minutes	123	25.5%

2.3 Data Normalization

The raw penalty kick data was processed to ensure consistency with prior research and to prepare it for analysis. In particular, for kicks taken by left-footed players, the directions were reversed so that shooting left always corresponds to the "natural" side[1] for all kickers.

[1] The natural side refers to the side of the goal that aligns with the kicker's dominant foot — typically the left side for right-footed players and the right side for left-footed players.

2.4 Comparison with Original Dataset

The dataset used in the original study was not publicly available and covered a significant earlier period (1997-2000). Our dataset, though more recent, offers both advantages and limitations. **Advantages:**

- Reflects modern football tactics, player behavior, and updated rules (e.g., VAR).
- Draws from a broader range of leagues and benefits from higher-quality, more accessible video sources.
- Temporal and contextual differences—such as changes in rules, playing styles, and player psychology—may limit direct comparability with the original 2002 study.

Limitations:

- Manual data collection may introduce inconsistencies or bias (same limitation as the original dataset)
- Lack of direct comparability due to differences in era, playing styles and rules mean findings may not be directly comparable to the 2002 study.

3 Framework

The game is modeled as a two-player, zero-sum game between a kicker and a goalie. Each player's strategies are directional choices: Left (L), Center (C), or Right (R). The payoffs represent the probability of scoring, depending on both the kicker's and goalie's choices. We reconstructed the payoff matrix based on the paper's notation and framework, where:

Kicker\Goalie	Left (L)	Center (C)	Right (R)
Left (L)	P_L	π_L	π_L
Center (C)	μ	0	μ
Right (R)	π_R	π_R	P_R

- π_L and π_R denote the scoring probabilities when the kicker shoots to the left or right, respectively, and the goalie jumps to the opposite side.
- μ is the scoring probability when the kicker shoots to the center and the goalie jumps to one of the sides (i.e., not center).
- P_L and P_R denote the scoring probabilities when the kicker shoots to the left or right, respectively, and the goalie jumps in the same direction.

3.1 Validating Theoretical Assumptions of the Model

We evaluate whether the data from our replication supports three core assumptions proposed in the original study: Sides and Center (SC), Natural Side (NS), and Kicker's Side (KS). These assumptions shape the structure of the zero-sum game between the kicker and goalie and can be tested using observed scoring probabilities. Figure 1 presents a side-by-side comparison of the results from the original paper (Table 1.A) and our replication (Table 1.B).

3.1.1 Assumption SC (Sides and Center). This assumption states that kickers are more likely to score when they avoid the direction the goalie chooses, and that shooting to one side yields higher success than shooting to the center when the goalie dives to that same side. Formally:

$$\pi_R > P_L, \quad \pi_L > P_R$$
$$\pi_R > \mu, \quad \pi_L > \mu$$

Comparison: Both datasets support this assumption. In our data, the scoring probability is significantly higher when the goalie guesses incorrectly (96.7% for natural side, 94.9% for opposite side) than when he guesses correctly (66.8% and 64.3%, respectively), reinforcing the importance of misdirection.

3.1.2 Assumption NS (Natural Side). This assumption states that kickers perform better when shooting to their natural side (e.g., right-footed players to the left).

Formally:

$$\pi_L \geq \pi_R, \quad P_L \geq P_R$$

Comparison: Both datasets confirm this. In our replication, success rates on the natural side are consistently higher than the opposite side, both when the goalie guesses right (66.8% vs. 64.3%) and when he guesses wrong (96.7% vs. 94.9%).

3.1.3 Assumption KS (Kicker's Side). KS refines NS by stating that the natural side is not only more successful overall, but also harder to save when the goalie guesses correctly.

Formally:

$$\pi_R - P_R \geq \pi_L - P_L$$

Comparison: The original paper found this assumption to hold, and our results confirm the same direction of effect. The scoring gap between correct and incorrect guesses is slightly smaller on the natural side than the opposite, indicating that natural-side shots are relatively harder to save even when anticipated.

4 Reproduction of the Empirical Tests

4.1 A. Testing the Assumption That Kickers and Goalies Move Simultaneously

4.1.1 Theoretical Framework. A fundamental assumption in game-theoretical models of penalty kicks is that the kicker and goalie move simultaneously, with neither of the players observing the other's action before committing to their own. If this assumption does not hold, strategic predictions of standard mixed-strategy equilibrium models may break down, as one player could condition their action on the other's move.

4.1.2 Hypothesis. We formally test:

- $H_0 : \beta = 0$ (Simultaneity holds; goalie's current action has no predictive power).
- $H_1 : \beta \neq 0$ (Simultaneity fails; kickers react to goalies).

4.1.3 Empirical Test. To test this assumption, we replicate the empirical approach from the original paper. Specifically, we examine whether the goalie's observed jump direction in a given penalty kick predicts the kicker's choice, controlling for historical tendencies of both players. If decisions are made simultaneously, the goalie's current jump direction should have no predictive power for the kicker's action on the same kick.

Table 1.A

	Goalie	
Kicker	Correct side	Middle or wrong side
Natural side ("left")	63.6 percent	94.4 percent
Opposite side ("right")	43.7 percent	89.3 percent

Notes: Table of the original paper.

Table 1.B

	Goalie	
Kicker	Correct side	Middle or wrong side
Natural side ("left")	66.8 percent	96.7 percent
Opposite side ("right")	64.3 percent	94.9 percent

Notes: Table of our replication.

Figure 1: Comparison of Table 1.A and Table 1.B: Observed Scoring Probabilities, by Foot and Side

We estimate a linear probability model where the dependent variable is an indicator for whether the kicker shoots to the right[2]:

$$R_i^K = X_i\alpha + \beta R_i^G + \gamma \overline{R}_i^K + \delta \overline{R}_i^G + \epsilon_i \quad (1)$$

Key Variables.

- **Dependent variable:** R_i^K — Indicator equal to 1 if the kicker shoots to the right, 0 otherwise.
- **Test variable:** R_i^G — Indicator equal to 1 if the goalie jumps to the right.
  The coefficient β on this variable tests the simultaneity assumption:
  - $\beta = 0$ supports the hypothesis that players move simultaneously.
  - $\beta \neq 0$ indicates a potential violation of simultaneity.
- **Controls:**
  - $\overline{R}_i^K$: Kicker's historical proportion of shots to the right.[3]
  - $\overline{R}_i^G$: goalie's historical proportion of jumps to the right.
  - X_i: Vector of game context controls — includes time of the game, scoreline, home team indicator, and league-year fixed effects.

Under the null hypothesis of simultaneous moves, the coefficient β should be zero.

4.1.4 Replication Results and Interpretation of Non-Simultaneity. Our replication reveals a statistically significant negative relationship between the goalie's jump direction and the kicker's choice: the coefficient on keeper jumps right is $\beta = -0.125$ ($p < 0.05$) across all specifications (Table 2.B, represented in Figure 3). This means that when the goalie jumps right, the probability the kicker shoots right decreases by 12.5 percentage. This sharply contrasts with the original study (Table 2.A, represented in Figure 3), where $\beta \approx 0$ (coefficients ranged from −0.026 to +0.027, all $p > 0.10$), supporting the simultaneity assumption. The divergence is noticeable in our bar chart plot comparing the coefficients of our replication with the original study represented in Figure 2. This contrast implies:

- **Violation of simultaneity:** The kicker may observe (or anticipate) the goalie's direction and react, contradicting the simultaneous-move assumption

The effect strengthens ($\beta = -0.173$) when analyzing only experienced kickers (≥ 4 penalty kicks), represented in columns 2 and 4 (Table 2.B), suggesting the violation is particularly pronounced among players with greater penalty-taking experience

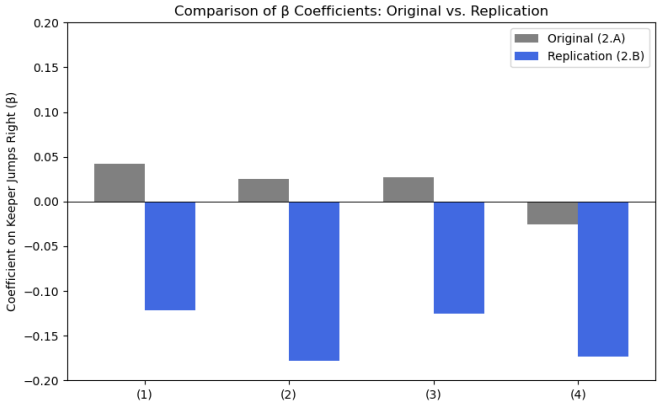

Figure 2: Comparison of β coefficients between the original study and replication. The graph shows the estimated effect sizes (with confidence intervals) for the variable "Keeper jumps Right." Numbers (1)-(4) represents the columns in Figure 3

4.1.5 Discussion and Implications. One likely explanation for this discrepancy is the difference in datasets. While the original study used data from the early 2000s, our analysis is based on a more recent dataset (post-2020). This temporal gap is crucial, as it may reflect structural changes in player behavior, training, and technology over the past two decades.

- **Strategic evolution:** Modern players may adopt more advanced strategies, including delayed run-ups or hesitation moves designed to read the opponent's action. For instance, top penalty kickers like Jorginho and Bruno Fernandes [2] frequently use stutter-step or hop techniques to observe the goalie's movement before committing to a direction.[4]

4.1.6 Conclusion. Our replication rejects the simultaneity assumption ($H_0 : \beta = 0$), finding that some kickers systematically react to goalies' observed jumps. Unlike the original study, which found no evidence that kickers respond to goalies' current actions, our results demonstrate that kickers do react to the goalie's jump direction, particularly among more experienced players. This does

[2]As a robustness check, we replicated our analysis using a probit model. The results were qualitatively similar to our linear probability model estimates, with no meaningful differences in either the significance or direction of the key coefficients.

[3]For both $\overline{R}_i^K$ and $\overline{R}_i^G$, we computed historical proportions based only on penalty kicks taken or faced *prior* to the one at stake. We verified that using alternative windows did not materially affect the results.

[4]These techniques aim to exploit timing asymmetries and are often effective in reducing goalie anticipation, thereby challenging the simultaneity assumption.

not imply that all players react in this way and in some penalty kicks the simultaneity assumption can still hold. This shift likely reflects evolving player strategies and advancements in training techniques that enable kickers to better observe and adapt to goalie behavior in real time.

4.2 B. Testing the Predictions of the Model That Are Robust to Aggregation

4.2.1 Theoretical Framework. Several key predictions should hold even when aggregating across heterogeneous kicker–goalie pairs. These aggregated predictions are derived from the structure of mixed-strategy equilibrium and remain valid without needing individual-level panel data. Below we summarize and test the five primary robust predictions:

(1) **Kickers should play "center" more often than goalies.**
(2) **Goalies should dive to the kicker's natural side (left) more frequently than kickers kick to that side.**
(3) **Under Assumption NS (Natural Side), goalies should dive left more often than right.**
(4) **Under Assumption KS (Kicker's Side), kickers should shoot left more often than right.**
(5) **The (Left, Left) action profile should be more common than (Left, Right) or (Right, Left), and those in turn more common than (Right, Right).**

4.2.2 Empirical Findings. Table 3.A from the original paper and our replicated Table 3.B, represented in Figure 4, both support the above predictions:

- In both datasets, kickers choose the center (*Middle*) more frequently than goalies do: 79 vs. 11 (original) and 54 vs. 12 (replication).
- Goalies dive left more often than kickers shoot left: 260 vs. 206 (original) and 265 vs. 247 (replication).
- Goalies prefer left to right: 260 vs. 188 (original) and 265 vs. 205 (replication).
- Kickers also favor left over right: 206 vs. 174 (original) and 247 vs. 181 (replication).
- The most frequent joint outcome is (Left, Left): 117 cases in original, 126 in our replication — higher than any other cell.

Thus, our replicated dataset confirms all five aggregated predictions of the original model.

4.2.3 Outcomes. Table 4.A and Table 4.B,represented in figure 5 present scoring probabilities across strategies. In both datasets, kickers appear to score more often when shooting to the center (especially when the goalie dives), which aligns with the theoretical expectation of *selection bias* in observed outcomes. Specifically:

- In the original data (Table 4.A), scoring is highest for central kicks (81.0%) and lowest when kickers shoot right (70.1%).
- In our replication (Table 4.B), this pattern persists with higher overall scoring rates (90.7% for center, 83.6% for right).

These differences in scoring probabilities across directions suggest that center kicks are selectively used by stronger players or in favorable situations — consistent with equilibrium theory under heterogeneity.

4.2.4 Limitations. While the aggregated predictions are supported, we must take into account the following:

- **Low statistical power:** Particularly for goalies choosing center, the small sample size reduces the reliability of estimated frequencies and success rates.
- **Data sparsity:** Most kicker–goalie pairs appear only once or twice in the data, limiting the ability to control for unobserved heterogeneity.

4.2.5 Conclusion: Despite the noted limitations, our replication strongly confirms the original paper results and the five aggregated predictions derived from the mixed-strategy equilibrium model.

4.3 C. Identical goalies

4.3.1 Theoretical Framework. The theoretical model becomes significantly more testable under the assumption of **Identical goalies** (IG). This assumption states that all goalies behave in the same way and are equally effective in saving penalty kicks. Formally, it implies that the payoff matrix depends only on the kicker and not the identity of the opposing goalie.

4.3.2 Implications of the IG Assumption. Assuming goalie homogeneity allows to:

- Gather all penalty kicks by a given kicker across matches, treating them as draws from a single distribution.
- Conduct player-specific tests of equilibrium conditions, such as equal scoring probability across strategies.
- Eliminate unobservable goalie-specific effects, simplifying identification and estimation.

This assumption is essential in both the original paper and in our replication, as most kicker–goalie pairs appear only once or twice in the dataset.

4.3.3 Empirical Evaluation: Are goalies Identical? In figure 6 tables **6.A** (original dataset) and **6.B** (replication) showcase F-tests on goalie fixed effects across four outcomes: scoring success, kicker shoots right, kicker shoots middle, and goalie jumps right.

- In Table 6.A, none of the F-tests reject the null hypothesis that goalie fixed effects are jointly zero. All p-values are above 0.19, suggesting no evidence of heterogeneity across goalies.
- Similarly, in Table 6.B, p-values range from 0.212 to 0.850, again indicating no statistically significant variation among goalies in the replication sample.

Importance of the F-statistic: The F-statistic tests whether including goalie fixed effects significantly improves model fit. A low F-statistic with a high p-value supports the IG assumption. In both datasets, the F-statistics are well below typical rejection thresholds, giving empirical support to the simplification that this assumption looks to give.

4.3.4 Testing Indifference Across Kicker Strategies. Given the IG assumption, we next test whether each kicker achieves close to equal

Table 3.A

Variable	(1)	(2)	(3)	(4)
Keeper jumps right	0.042	0.025	0.027	−0.026
	(0.052)	(0.062)	(0.052)	(0.060)
Kicker's % shots to the right, excluding this kick	0.219	0.370	0.220	0.357
	(0.082)	(0.122)	(0.082)	(0.126)
Goalie's % jumps to the right, excluding this kick	−0.032	0.001	0.012	0.001
	(0.103)	(0.131)	(0.104)	(0.135)
(League × year) dummies included?	yes	yes	yes	yes
Full set of covariates included?	no	yes	no	yes
Sample limited to kickers with 4+ kicks?	no	no	yes	yes
R^2	0.029	0.051	0.068	0.087
Number of observations	373	252	373	252

Notes: Original paper table

Table 3.B

Variable	(1)	(2)	(3)	(4)
Keeper jumps right	−0.122	−0.178	−0.125	−0.173
	(0.051)	(0.061)	(0.052)	(0.064)
Kicker's % shots to the right, excluding this kick	0.034	-0.128	0.033	-0.087
	(0.088)	(0.153)	(0.089)	(0.164)
Goalie's % jumps to the right, excluding this kick	−0.09	−0.02	−0.092	0.001
	(0.103)	(0.131)	(0.108)	(0.138)
(League × year) dummies included?	yes	yes	yes	yes
Full set of covariates included?	no	yes	no	yes
Sample limited to kickers with 4+ kicks?	no	no	yes	yes
R^2	0.02	0.043	0.052	0.092
Number of observations	360	254	360	254

Notes: Table with our dataset. Only for kickers and goalies with $\geq$ 2 penalty kicks

Figure 3: Comparison of Table 3.A and Table 3.B: Testing the Assumption that the Kicker and Goalie Move Simultaneously (Dependent Variable: Kicker Shoots Right)

Table 4.A

	Kicker			
Goalie	Left	Middle	Right	Total
Left	117	48	95	260
Middle	4	3	4	11
Right	85	28	75	188
Total	206	79	174	459

Notes: Original paper Table 3.

Table 4.B

	Kicker			
Goalie	Left	Middle	Right	Total
Left	126	30	109	265
Middle	5	2	5	12
Right	116	22	67	205
Total	247	54	181	482

Notes: Dataset used in this replication

Figure 4: Comparison of the tables 4.A and 4.B with the Observed Matrix of Shots Taken

Table 5.A

	Kicker			
Goalie	Left	Middle	Right	Total
Left	63.2	81.2	89.5	76.2
Middle	100	0	100	72.7
Right	94.1	89.3	44.0	73.4
Total	76.7	81.0	70.1	74.9

Notes: Original paper Table 4.

Table 5.B

	Kicker			
Goalie	Left	Middle	Right	Total
Left	66.8	96.7	94.6	81.7
Middle	100.0	0.0	100.0	83.3
Right	96.6	90.9	64.3	85.4
Total	81.5	90.7	83.6	83.3

Notes: Dataset used in this replication

Figure 5: Comparison of tables 5.A and 5.B with the Observed Matrices of Outcomes: Percentage of Shots in Which a Goal Is Scored

expected payoff across the directions they choose: left, center, or right. This is a central condition for the mixed-strategy equilibrium.

In figure 7 tables **7.A** (original dataset) and **7.B** (replication) report F-tests on whether, for each kicker, the probability of scoring is equal across the chosen strategies.

- In Table 7.A, the null hypothesis is rejected at the 10% level for 5 out of 27 kickers with at least five kicks, and 4 out of 9 kickers with at least eight kicks.
- In contrast, Table 7.B shows much weaker evidence of rejection: none of the 30 kickers (with five or more kicks) or 12 kickers (with eight or more) show significant deviations from equality of scoring probabilities. All p-values are well above 0.90.

4.3.5 Possible Reasons for Differences Between Tables A and B. Despite similar methodologies our results boost the original paper conclusions, there are several plausible reasons for this:

- **Temporal effects:** The original paper's data is from 1997–2000 (French and Italian leagues), while our replication includes more recent seasons: English Premier League (2023–24), Portuguese League (2024–25), and Italian League (2020–21, 2022–23). Over time, the tactical training and use of data analytics in football have significantly evolved, potentially increasing strategic randomness and reducing exploitable patterns.
- **League-specific behavior:** Different leagues may exhibit distinct penalty-taking cultures or goalie training styles. For instance, the Premier League and Portuguese League may

Table 6.A

Independent variable	Kick successful	Kicker shoots right	Kicker shoots middle	Goalie jumps right
F statistic	0.95	0.98	0.88	1.14
$[p = \ldots]$	[0.571]	[0.52]	[0.70]	[0.19]
Coefficients on other covariates:				
Minute 0–14	0.512 (0.134)	−0.220 (0.144)	0.113 (0.109)	0.150 (0.154)
Minute 15–29	0.219 (0.121)	−0.194 (0.140)	0.049 (0.104)	0.104 (0.150)
Minute 30–44	0.254 (0.120)	0.038 (0.139)	0.003 (0.104)	0.080 (0.149)
Minute 45–59	0.124 (0.107)	0.115 (0.112)	0.009 (0.093)	0.095 (0.119)
Minute 60–74	0.112 (0.105)	0.113 (0.113)	0.003 (0.093)	0.089 (0.117)
(League × year) dummies?	yes	yes	yes	yes
Kicker FE?	yes	yes	yes	yes
Goalie FE?	yes	yes	yes	yes
R^2	0.552	0.571	0.532	0.557

Table 6.B

Independent variable	Kick successful	Kicker shoots right	Kicker shoots middle	Goalie jumps right
F statistic	1.192	0.772	0.881	0.801
$[p = \ldots]$	[0.212]	[0.850]	[0.689]	[0.813]
Coefficients on other covariates:				
Minute 0–14	−0.032 (0.108)	0.152 (0.166)	0.045 (0.109)	−0.088 (0.171)
Minute 15–29	−0.032 (0.084)	−0.069 (0.130)	0.029 (0.086)	−0.111 (0.134)
Minute 30–44	0.058 (0.075)	0.153 (0.116)	0.081 (0.077)	0.058 (0.120)
Minute 45–59	0.061 (0.089)	0.182 (0.137)	−0.014 (0.090)	−0.089 (0.141)
Minute 60–74	0.088 (0.086)	−0.057 (0.132)	0.165 (0.087)	−0.033 (0.136)
(League × year) dummies?	yes	yes	yes	yes
Kicker FE?	yes	yes	yes	yes
Goalie FE?	yes	yes	yes	yes
R^2	0.701	0.571	0.596	0.574

Notes: Table 6.A is from the original paper. Table 6.B is from our replication (goalies with ≥ 2 kicks).

Figure 6: Comparison of tables 6.A and 6.B that test whether the Goalies are Homogeneous

emphasize statistical preparation, making player behavior more equilibrium-consistent compared to late 1990s Serie A or Ligue 1.

4.3.6 Critical Perspective.

Is the IG Assumption Realistic? Although not rejected statistically, the IG assumption simplifies away real-world variation in goalie ability and behavior. Differences in anticipation, reach are well known among professionals. However, due to limited repeated observations per goalie, estimating individual goalie effects reliably remains difficult in practice.

4.3.7 Limitations. While the replication shows few rejections, statistical power remains a concern, especially for individual-level equilibrium tests with small samples (5–10 kicks per player). The inability to reject may reflect insufficient data, not necessarily indifference.

4.3.8 Conclusion. Despite its simplifications, the IG assumption is supported by the data and enables meaningful within-kicker testing of equilibrium predictions. Our replication not only confirms the homogeneity of goalie behavior but also finds strong evidence that kickers achieve approximate scoring indifference across chosen strategies, a key observation of mixed-strategy equilibrium.

5 Conclusion

This replication of "Testing Mixed-Strategy Equilibria When Players Are Heterogeneous: The Case of Penalty Kicks in Soccer" [1] aimed to test whether modern footballers with different characteristics adhere to mixed-strategy equilibria in penalty kicks. Using a newly

Table 7.A

	Kickers with five or more kicks		Kickers with eight or more kicks	
	(1)	(2)	(3)	(4)
A. Null hypothesis: For a given kicker, the probability of scoring is the same when kicking right, middle, or left.				
P value of joint test	0.10	0.28	0.15	0.45
F statistic	1.36	1.15	1.14	1.01
Degrees of freedom (numerator, denominator)	43:136	43:123	16:76	16:63
Number of individual kickers	27	27	9	9
Number for whom null is rejected at 0.10	5	5	4	1
Full set of covariates included?	no	yes	no	yes
B. Null hypothesis: For goalies facing a given kicker, the probability of scoring is the same whether the goalie jumps right or left				
P value of joint test	0.31	0.28	0.42	0.19
F statistic	1.14	1.16	1.04	1.45
Degrees of freedom (numerator, denominator)	27:146	27:133	9:80	9:67
Number of individual kickers	27	27	9	9
Number for whom null is rejected at 0.10	5	4	4	1
Full set of covariates included?	no	yes	no	yes

Notes: Table of the original paper. Statistics are based on linear probability models in which the dependent variable is whether or not a goal is scored. See text for details.

Table 7.B

	Kickers with five or more kicks		Kickers with eight or more kicks	
	(1)	(2)	(3)	(4)
A. Null hypothesis: For a given kicker, the probability of scoring is the same when kicking right, middle, or left.				
P value of joint test	0.97	0.98	0.99	0.95
F statistic	0.64	0.61	0.44	0.55
Degrees of freedom (numerator, denominator)	58:136	58:123	24:86	24:73
Number of individual kickers	30	30	12	12
Number for whom null is rejected at 0.10	0	0	0	0
Full set of covariates included?	no	yes	no	yes
B. Null hypothesis: For goalies facing a given kicker, the probability of scoring is the same whether the goalie jumps right or left				
P value of joint test	0.26	0.18	0.70	0.73
F statistic	1.18	1.26	0.75	0.72
Degrees of freedom (numerator, denominator)	30:164	30:151	12:98	12:85
Number of individual kickers	30	30	12	12
Number for whom null is rejected at 0.10	3	4	1	3
Full set of covariates included?	no	yes	no	yes

Notes: Table of our replication. Statistics are based on linear probability models in which the dependent variable is whether or not a goal is scored. See text for details.

Figure 7: Comparison of Table 7.A and Table 7.B: Testing for Identical Scoring Probabilities Across Left, Middle, and Right for Individual Kickers and the Goalies They Face

compiled dataset from three major European leagues spanning 2020–2025, we revisited the core assumptions and empirical tests of the original paper.

Our findings offer mixed but insightful results. First, we partially reject the simultaneity assumption: unlike in the original dataset, modern kickers in particular experienced ones appear to react to the goalie's actions. This likely reflects strategic evolution, such as the use of delayed run-ups.

Second, the five theoretical predictions that are robust to aggregation were confirmed in our data. Kickers and goalies continue to show predictable patterns in their choices and scoring probabilities consistent with equilibrium play.

Third, under the Identical goalies (IG) assumption, our replication supports equilibrium behavior: we find no evidence that scoring probabilities differ across directions for individual kickers.

Despite differences in era, league, and data collection methods, our replication broadly supports the validity of the original theoretical framework when adapted to the modern game. At the same time, our results suggest that strategy execution in football has become more dynamic and responsive, opening new questions about how evolving tactics interact with game-theoretic predictions.

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] Twelve Yards. 2024. *Bruno Fernandes' Penalty Secret.* https://twelveyards.substack.com/p/bruno-fernandess-penalty-secret