Professionals Play Minimax: Appendix

Ignacio Palacios-Huerta
Brown University

July 2002

In this supplementary appendix we first report additional descriptive statistics in section 1. We begin by reporting the empirical payoff matrix for games involving left-footed kickers and right-footed kickers on a separate basis, as well as the Nash equilibria and the observed empirical frequencies for these two games. We then report the distribution of strategies and scoring rates over time in the dataset. Among other things, we observe that the scoring rate is lower for penalties shot in the last ten minutes of a match, especially in close games. We conjecture that this lower rate may be attributed to "nervousness." We examine in section 2 the role that nervousness may play on the tests of equality of winning rates across strategies. In section 3 we provide evidence that supports the idea that for a given player one given game is being played against randomly selected opponents. In section 4 we evaluate the individual and aggregate tests of equality of winning rates across strategies as a three-action game for a limited number of players. Lastly, in section 5 we examine whether professional players also generate sequences with no serial correlation in penalty shootouts where penalties are taken in rapid succession in a short span of time.

1. Descriptive Statistics

1.1. Symmetry Across Types of Kickers

The empirical payoff matrices for left-footed and right-footed kickers when considered on a separate basis are reported in Table 1A, along with the mixed strategy Nash equilibria and the observed empirical frequencies for these two groups of players.

[Table 1A Here]

It can be readily observed that empirical behavior is remarkably close to the Nash predicted frequencies for each group. This is consistent with the findings at the aggregate level, and with the fact that the hypothesis that these two games are identical up to the renaming of the actions cannot be rejected.[1]

1.2. Strategic Choices and Scoring Rates Over Time

Soccer matches last two equal periods of 45 minutes, with a 15-minute half-time interval.[2] Table 2A shows the distribution of strategies and scoring rates by halves, in the last 10 minutes of the match, and by score differences.

[Table 2A Here]

Most penalty kicks take place in the second half, where a relatively large amount takes place in the last ten minutes. It may be noted that the scoring rate appears to decrease over time: from 82.9 in the first half to 78.3 percent in the second half. The main reason for this decrease is the notably lower scoring rate in the last ten minutes of the game: 73.3 percent. This is the case both at the aggregate level for all penalty kicks and for given score difference. Across score differences, it initially appears that the scoring rate is slightly greater the closer the game is at the time of the penalty shot (that is, it is greatest in ties, followed by the rates at the '-1' and '1' score differences, and then at '-2' and '2'). This is the case when penalties are sorted by halves. However, the opposite impression arises when we examine the penalties shot in the last ten minutes of the match. In these cases, the closer the game is, and the more important an additional goal would seem to be for the kicker's team, the *lower* the scoring rate appears to be. The scoring rate is lowest when games are tied and when the kicker's team is behind by one goal. This phenomenon may be interpreted as "nervousness" in "big penalties," and is often observed in other sports as well. For instance, in professional NBA basketball, free throw shooters in the last minute in close playoff games have a notably lower scoring rate than in regular season

[1] See Palacios-Huerta (2002). As is also indicated there, the power functions associated with the aggregate Pearson tests for left-footed kickers, right-footed kickers, and for the goalkeepers facing the two types of kickers are virtually identical to those presented in Palacios-Huerta (2002, Figure 2).

[2] Allowance for time lost is made in either period at the discretion of the referee for all time lost through substitutions, removal of injuried players, wasting time or any other cause. We consider the penalties shot in this extra time as shot in the last minute of each half.

games that are not close, even though in this play individuals face no opponents and have no strategic choices to make.[3] Next, we examine whether "nervousness" may have a noticeable impact on the tests of equality of winning rates across strategies for the different players.[4]

2. Nervousness and the Tests of Equality of Winning Rates

As indicated above, the descriptive evidence in the previous table allows us to conjecture that nervousness may be a potential determinant of the scoring rate. This conjecture is intuitive and consistent with the fact that many superstars have missed or had penalties saved in critical situations of stress and pressure of great magnitude (Miller (1998)).[5] It is also consistent with the evidence from other plays at the professional level, even when players face no opponents and have no strategic choices to make (e.g., free throws in professional basketball). However, nervousness in itself need not have any effect on the tests of equality of scoring probabilities, unless the choice of strategy is related to the importance of the penalty.

First, a logit regression (not reported) for the choice of natural side that includes all the independent variables considered in Table 6 in Palacios-Huerta (2002) shows a positive but clearly insignificant effect on "important" penalties (those penalties shot in the last ten minutes of a match when the score difference was 0, 1, or -1). Second, in order to evaluate the extent of the effects in the tests of equality of scoring probabilities at the individual and aggregate levels, we consider the same subsample of kickers and goalkeepers *except* these important penalties. These observations represent 12.9 percent of the sample. We then perform the same individual and aggregate statistical tests as in Tables 3 and 4 in Palacios-Huerta (2002). The results are shown in Table 3A.

[Table 3A Here]

[3] See for instance the ESPN Magazine, January-June 2000 issues. Other examples where players face no opponents, have no strategic choices to make, and where winning rates are notably lower in "very important" situations include archery and skeet shooting.

[4] In soccer, teams are free to choose the penalty kicker. The data also include some evidence on the location of the shot (its height). However, there is little variation across penalties and location plays no noticeable role in the dataset.

[5] As is mentioned in Palacios-Huerta (2002), a logit regression for the scoring rate confirms this conjecture for the penalties shot in the last ten minutes of close games (when the score difference is 0,1, or -1). Nervousness, rather than fatigue, is the typical reason given by professional players in interviews (Miller (1998)), an interpretation that is consistent with the low level of effort that needs to be supplied in a penalty kick.

The null hypothesis is rejected for 2 players at the 5 percent level and for 4 players at the 10 percent level. Recall that 2.1 and 4.2 are the number of rejections we would expect from a sample with forty-two experiments, and that 3 and 5 rejections, respectively, were obtained in Palacios-Huerta (2002, Table 3). The results indicate that, even though the number of rejections in the full sample were already basically those predicted by the theory, the number of rejections now virtually coincides with the predictions. Note also that the p-values of the Pearson statistics continue to be distributed quite uniformly across deciles.

With regard to the tests at the aggregate level, the Pearson statistic of the joint test of the null hypothesis that all observations are generated by equilibrium play is 38.836, with a p-value of .610. This result is similar to that in the full sample. Likewise, similar results are obtained for goalkeepers and kickers as separate groups. The results of the KS tests still decisively support the hypothesis of equality of winning rates, although the p-values are slightly lower than in the full sample.[6]

We conclude from the analysis that the results of the tests of equality of scoring rates at individual and aggregate levels continue to support quite strongly the first implication of the Minimax Theorem when the most important penalties are considered on a separate basis.

3. Tests of Equality of Winning Rates for Subsample Tests

An interpretation in the literature of the mixed strategy equilibrium is that it provides a good description of the steady state behavior of players who play *one* given game repeatedly against randomly selected opponents (Osborne and Rubinstein, 1994, pp. 38–39). A simple test could then detect whether in fact more than one game was actually being played. The idea is that if in fact two or more different games were played—but somehow the tests of equality of winning rates in the whole sample did not reject the hypothesis of equality of winning rates across strategies—then testing the hypothesis for different subsamples of the data would tend to reject the hypothesis that winning rates are identical, or at least show significant variation in the p-values of the test. This idea is implemented next for many different subsets, all of the same

[6]The power functions of these tests are essentially the same as those in Figure 2 in Palacios-Huerta (2002). If we assume that "representative" players participate in every penalty kick, the Pearson statistic is .032 (p-value: .858) for "important" penalties, and .797 (p-value: .371) for "non-important" penalties. Similar values are found for kickers alone and for goalkeepers alone for each class of penalties.

size, randomly chosen for each player. The first three columns in Table 4A report the sample size per subset, and the average p-value and standard deviation of the p-values of the Pearson's tests of the hypothesis of equality of winning rates when evaluated for 100 subsets for each player.

[Table 4A Here]

The results show that the average p-values are very similar to those found in Palacios-Huerta (2002, Table 3) for each and every kicker and goalkeeper in the sample. Moreover, the standard deviation of the p-values for a given player is relatively low, always below .187. This great stability in p-values helps confirm the idea that for a given player one given game is being played against randomly selected opponents.

Additional empirical support may also be obtained by testing the assumption that opponents are homogeneous (taking into account, as in Palacios-Huerta (2002), the natural sides for the players). This idea may be tested using a regression framework and implementing an F-test of the joint significance of opponents' fixed effects. If opponents are statistically homogeneous, the F-test should not reject the null hypothesis that all opponent-fixed effects are equal. This regression is implemented for various dependent variables (including the direction of the shot $\{L, R\}$ and whether a goal is scored or not), limiting the sample to opponents with at least 10 observations in the dataset, and including various covariates such as indicators corresponding to the different 10-minute segments of the game, the score at the time of the penalty kick, whether the kicker plays for the home team, and league and year interactions. The results of the F-tests are available upon request. They show that in 2 cases can we reject the hypothesis that opponents are statistically identical at the 5% level, and in 4 cases at the 10% level. These results are just those predicted by chance with 42 observations.[7] Hence, they again conclusively support the interpretation that for any given player one given game is being played against randomly selected opponents.

Lastly, additional evidence is also consistent with these results. If a given player plays one given game against various opponents as the results in Table 4A and the F-tests indicate, then the elements of the payoff matrix should be quite "stable." This idea may be evaluated using the technique of bootstrapping to obtain confidence intervals on the payoffs in each of the cells for given player by studying a sample

[7] Interestingly, p-values tend to be greater for goalkeepers than for kickers, indicating that goalkeepers may tend to be more homogenous than kickers.

of bootstrap estimators obtained by sampling m observations *with replacement*, and recomputing the payoffs with each sample. The number of bootstrap repetitions is chosen following Andrews and Buchinsky (2000). The results show that the sampling variability of the estimator of the payoffs is very small. In particular, the standard deviation for π_{LL} and π_{RR} is between 0.021 and 0.043 across all the different players in the sample, whereas for π_{LR} and π_{RL} it is below 0.015 in all cases. The variation in π_{LR} and π_{RL} is notably lower mostly because the scoring rate is always extremely high when kickers and goalkeepers strategies do not coincide. The stability in the payoffs for each and every player is again consistent with the idea that one game is being played against various opponents.

4. Testing for Equality of Winning Rates in a Three-Action Game

Most players in the sample, particularly goalkeepers, never choose C or choose it just once or twice over the 1995-2000 period we examine. Therefore, having so few or no observations for them, it is not possible to conduct individual level tests of the Minimax hypothesis when the strategic situations are treated as a three-action game. Moreover, as indicated in Palacios-Huerta (2002, footnote 11) and Chiappori et al (2000), the availability of C as an action is not a real issue in the empirical analysis of aggregate data across players. Yet, it may be of interest to evaluate the tests of winning rates across strategies in a three-action game for the small number of players with most C observations. There are 10 kickers that were categorized to have at least 3 kicks to the center. The tests are implemented in Table 5A for these players.

[Table 5A Here]

Interestingly, the results in Panel A are quite supportive of the hypothesis of equality of winning rates across the three actions. There is only 1 rejection at the 10 percent level, exactly the amount predicted with 10 players. Also, the p-values of the tests are distributed quite uniformly across deciles. Consistent with this, the Pearson and KS tests in Panel B at the aggregate level strongly support the hypothesis that the data for all individual experiments were generated by equilibrium play.

We conclude that these results support the null hypothesis of equal success rates across strategies for these players. As is well known, however, when the amount of observations for one of the strategies is low, a problem with the KS test is that it does not have substantial power against alternative hypotheses (see, for instance, Massey (1951)).

5. Testing for Serial Correlation in Penalty Shootouts

It is known that the equilibrium strategies in repeated zero-sum games are independent of the time lags between the stages of the game. Moreover, the equilibrium strategies in the repeated game dictate that at every stage players play according to the equilibrium strategy of the stage game. Therefore, from the theoretical perspective, time lags have no influence on equilibrium play. Yet, it is empirically of interest to evaluate whether time lags play a role in the result of serial independence obtained in the natural setting studied in Palacios-Huerta (2002). Recall that professional soccer players were involved in no more than 15-20 penalty shots in a given year. Here we test for serial correlation in penalty kick situations where repeated choices are made in rapid succession. Penalty shootouts are used in elimination tournaments to break ties. In the typical penalty shootout, five different players from each team shoot a penalty kick each to the goalkeeper of the opposing team in a short span of time. The two teams alternate taking one penalty at a time.[8] The data we examine below come from 82 penalty shootouts (817 observations) in games in the World Cup 1982-1998 and European Cup 1976-2000, and in the yearly Champions and UEFA Cups, Spanish Cup, Italian Cup, and English Cup during 1990-2000.

[Table 6A Here]

We first test in Panel A whether goalkeepers are homogeneous.[9] We use a regression framework for different outcome variables following the idea in Chiappori et al (2000). The null hypothesis that goalkeepers are homogeneous corresponds to the goalkeeper-fixed effects being jointly insignificantly different from zero in different outcome variables: a goal is scored, the kicker shoots to the natural side, and the goalkeeper goes to the natural side. The analysis includes as explanatory variables various covariates describing the state of the shootout (shootout tied, goalkeepers'

[8] If the tie persists after each team has kicked five penalty kicks, further penalties are shot until the tie is broken. Not all five penalties need to be shot if the winning team of the penalty shootout can be decided before all ten penalties are shot (e.g., if after each team has shot 3 penalties the score is 3-0 for one of the teams).

[9] It is clearly not possible to test for homogeneous kickers given that there is only one observation per kicker.

team ahead by one goal, goalkeepers' team is ahead by 2 or more goals, and goalkeepers' team behind by one goal, and goalkeeper is in the home team), as well as goalkeeper- and kicker-fixed effects. We find that in none of the cases can we reject the hypothesis that all goalkeepers are identical: the p-values of the F-statistic for the joint significance of goalkeeper-fixed effects are .57, .71, and .69.

Having established that goalkeepers are statistically homogeneous, we can then use *all* observations in the shootouts to estimate a logit equation for goalkeepers' choice of natural side. We then test for serial correlation in Panel B. The main result in this panel is that the null hypothesis that *all* the explanatory variables involving own and opponents' lagged strategies are jointly statistically insignificant (hypothesis #1) cannot be rejected at conventional significance levels. Likewise, hypothesis #2–#5 cannot be rejected at conventional significance levels either.[10]

We conclude from these findings that professional players generate sequences with no serial correlation in both natural settings, when strategies are spaced out over time and when strategies must be taken in rapid succession.

[10]The result in Panel A for goalkeepers is also obtained in Chiappori et al. (2000). Interestingly enough, even though it is not possible to test for homogeneous kickers, the results on randomization in Panel B are similar for the kickers when they are all treated as an "homogeneous" kicker; hypotheses #1–#5 cannot be rejected for them either.

References

Andrews, D. W. K. and Buchinsky, M. (2000), “A Three-Step Method for Choosing the Number of Bootstrap Repetitions”, *Econometrica,* 68, 23-51.

Chiappori, P. A., Levitt, S. D. and Groseclose, T. (2000): “Testing Mixed Strategy Equilibrium When Players Are Heterogeneous: The Case of Penalty Kicks”, *American Economic Review*, forthcoming.

Massey, F. J. (1951), “The Kolmogorov-Smirnov Test for Goodness of Fit”, *Journal of the American Statistical Association*, 46, 68-78.

Miller, C. (1998), *He Always Puts It to the Right. A Concise History of the Penalty Kick.* Victor Gollancz, St. Edmundsbury Press Ltd., London.

Osborne, M. J. and Rubinstein, A. (1994), *A Course in Game Theory.* MIT Press.

Palacios-Huerta, I. (2002), “Professional Play Minimax”, Mimeo, Department of Economics, Brown University.

TABLE 1A

Payoffs and Mixing Probabilities by Kicker Type

Right-Footed Kickers

PAYOFFS

		Goalkeeper *L*	Goalkeeper *R*
Kicker	*L*	57.23	95.35
	R	93.20	68.79

MIXING PROBABILITIES

	g_L	$1 - g_L$	k_L	$1 - k_L$
Nash Predicted	42.47%	57.53%	39.03%	60.97%
Actual	43.62%	56.38%	40.11%	59.98%

Left-Footed Kickers

PAYOFFS

		Goalkeeper *L*	Goalkeeper *R*
Kicker	*L*	60.46	94.20
	R	92.32	72.20

MIXING PROBABILITIES

	g_L	$1 - g_L$	k_L	$1 - k_L$
Nash Predicted	40.84%	59.16%	37.35%	62.65%
Actual	41.20%	58.80%	39.70%	60.30%

Note: In each game ¨R¨ denotes ¨kicker´s natural side¨ and ¨L¨ denotes ¨kicker´s ¨non-natural side.¨

TABLE 2A

Distribution of Strategies and Scoring Rates over Time

Score Difference	#Obs.	LL	LC	LR	CL	CC	CR	RL	RC	RR	Scoring Rate
						FIRST HALF					
0	307	18.9	1.3	16.9	4.2	0.9	5.2	27.0	0.6	23.7	83.0
1	95	18.9	0	21.0	4.2	0	3.1	27.3	0	25.2	78.9
-1	111	21.6	0.9	25.2	1.8	0	0.9	19.8	0.9	28.8	83.7
2	28	25.0	3.5	21.4	3.5	0	0	21.4	0	28.5	78.5
-2	5	40.0	0	20.0	0	0	0	20.0	0	20.0	80.0
All Penalties	558	21.1	0.8	19.8	3.9	0.3	3.5	20.0	0.3	29.7	82.2
						SECOND HALF					
0	273	14.6	1.4	25.6	4.3	0.7	6.2	14.6	0.7	32.6	80.5
1	140	19.2	0	17.8	4.2	0	2.1	28.5	0	27.8	77.1
-1	203	18.7	0.9	26.1	1.9	0	0.4	20.1	0.4	31.0	79.8
2	69	23.1	1.4	15.9	5.7	0	0	20.2	1.0	30.4	73.9
-2	109	25.6	0	25.6	3.6	0	3.6	16.5	0	24.7	78.0
All Penalties	859	18.7	0.9	23.2	3.3	0.3	3.6	22.8	0.5	26.3	78.3
						LAST 10 MINUTES					
0	80	20.0	1.2	22.5	0	0	3.7	20.0	0	32.5	70.0
1	26	26.9	0	19.2	0	0	0	23.0	0	28.5	73.0
-1	87	24.1	0	18.3	1.1	0	0	22.9	0	34.4	70.1
2	16	18.7	0	18.7	0.6	0	0	25.0	0	31.2	81.2
-2	37	24.3	0	24.3	0	0	2.7	21.6	0	27.0	81.0
All Penalties	266	21.8	0.4	21.0	0.3	0	0.7	25.1	0	30.8	73.3

TABLE 3A

Tests for Equality of Scoring Probabilities Excluding Close Games in Last 10 Minutes

Kicker	Obs.	Pearson Statistic	p-value
1.	30	0.000	0.999
2.	31	0.020	0.902
3.	35	0.110	0.735
4.	37	0.083	0.773
5.	35	0.054	0.817
6.	32	0.050	0.823
7.	40	0.156	0.693
8.	35	0.341	0.559
9.	27	0.241	0.623
10.	32	0.163	0.687
11.	30	0.536	0.464
12.	32	0.521	0.471
13.	34	0.344	0.558
14.	28	1.207	0.272
15.	26	0.337	0.539
16.	37	0.757	0.384
17.	32	3.752	0.053*
18.	46	1.787	0.182
19.	39	2.503	0.114
20.	36	1.496	0.221
21.	33	2.432	0.210
22.	34	4.545	0.033**

Kicker	Obs.	Pearson Statistic	p-value
1.	35	0.030	0.863
2.	32	0.163	0.687
3.	26	0.001	0.973
4.	42	0.064	0.800
5.	31	0.880	0.348
6.	34	0.147	0.702
7.	33	0.188	0.665
8.	31	0.420	0.588
9.	28	0.608	0.436
10.	35	0.033	0.855
11.	30	0.384	0.536
12.	30	0.545	0.460
13.	30	0.305	0.580
14.	39	0.507	0.476
15.	40	0.925	0.336
16.	30	3.000	0.083*
17.	36	1.575	0.209
18.	37	4.359	0.037**
19.	38	1.121	0.290
20.	35	2.146	0.702

	Pearson Test		Kolmogorov-Smirnov Test	
	Statistic	p-value	Statistic	p-value
All players	38.836	0.610	0.643	0.792
All kickers	21.435	0.494	0.467	0.802
All goalkeepers	17.401	0.626	0.417	0.712

TABLE 4A

Tests for Equality of Scoring Probabilities for Subsample Sets

Kicker	Total Obs.	Obs.per Subset	Pearson Tests Avg. *p*-value	Pearson Tests Std. dev. *p*-value
1.	34	28	0.970	0.021
2.	41	34	0.902	0.072
3.	40	32	0.855	0.103
4.	38	32	0.735	0.111
5.	38	32	0.676	0.155
6.	36	30	0.667	0.098
7.	41	34	0.662	0.152
8.	35	28	0.656	0.167
9.	31	25	0.519	0.185
10.	35	28	0.490	0.164
11.	32	25	0.471	0.187
12.	32	25	0.471	0.149
13.	38	30	0.341	0.163
14.	30	25	0.333	0.137
15.	30	25	0.283	0.107
16.	42	35	0.257	0.082
17.	40	34	0.251	0.130
18.	46	27	0.201	0.091
19.	39	32	0.197	0.102
20.	40	34	0.088*	0.067
21.	40	34	0.038**	0.027
22.	40	34	0.030**	0.022

Kicker	Total obs.	Obs.per subset	Pearson Tests Avg. *p*-value	Pearson Tests Std. dev. *p*-value
1.	37	31	0.982	0.102
2.	38	32	0.898	0.060
3.	30	25	0.866	0.101
4.	50	42	0.804	0.098
5.	36	28	0.777	0.155
6.	34	28	0.702	0.147
7.	37	31	0.638	0.109
8.	37	31	0.588	0.176
9.	32	25	0.568	0.180
10.	40	34	0.533	0.172
11.	33	26	0.519	0.163
12.	30	25	0.460	0.180
13.	34	28	0.447	0.144
14.	40	34	0.429	0.153
15.	44	37	0.328	0.082
16.	36	29	0.298	0.121
17.	42	36	0.118	0.071
18.	42	36	0.113	0.082
19.	42	36	0.071*	0.062
20.	40	34	0.024**	0.022

*Note: This table reports for each player the average *p*-value and the standard deviation of the *p*-values of one hundred tests of the equality of winning probabilities across strategies using randomly selected subsamples of all his observations with replacement.

Table 5A

Tests for Equality of Scoring Probabilities

Panel A: Pearson Tests for Individual Players in 3-Action Game

Kicker	Obs.	Mixture L	Mixture R	Mixture C	Scoring Rate L	Scoring Rate R	Scoring Rate C	Pearson statistic	Pearson *p*-value
1. Mendieta (Lazio)	34	0.32	0.59	0.09	0.90	0.85	1.00	0.677	0.712
3. Zidane (Real Madrid)	40	0.45	0.47	0.07	0.83	0.84	1.00	0.577	0.749
5. Mihajlovic (Lazio)	38	0.42	0.45	0.13	0.75	0.82	0.80	0.271	0.872
6. Signori (Bolonia)	36	0.27	0.61	0.11	0.70	0.77	0.75	0.193	0.907
7. Rui Costa (AC Milan)	41	0.19	0.71	0.10	0.62	0.86	1.00	3.401	0.182*
8. Amoroso (Udinese)	35	0.32	0.57	0.11	0.72	0.85	0.75	0.738	0.691
11. Del Piero (Juventus)	32	0.44	0.44	0.12	0.78	0.92	0.75	1.388	0.499
14. Garitano (Zaragoza)	30	0.33	0.53	0.13	0.90	0.75	0.75	0.937	0.625
20. Donosti (Eibar)	40	0.35	0.57	0.07	0.92	0.65	1.00	4.754	0.092**
22. Penev (At. Madrid)	40	0.40	0.50	0.10	0.93	0.85	0.75	1.257	0.533
All kickers	366	0.349	0.546	0.103	0.820	0.815	0.842		

Panel B: Tests for Equality of Scoring Probabilities for Aggregate Distributions

The Pearson test evaluates the hypothesis that the data for all experiments were generated by equilibrium play: $p^i{}_L = p^i{}_R = p^i{}_C$ for each player *i,* allowing for differences in probabilities across players. The Kolmogorov-Smirnov test evaluates the null hypothesis that the empirical distribution of *p*-values in individual Pearson tests was generated by random draws from the Uniform distribution $U[0,1]$.

	Statistic	*p*-value
Pearson test	14.19	0.820
Kolmogorov-Smirnov test	0.199	0.501

Note: *Indicates rejection at the 10 percent level, and ** indicates rejection at the 20 percent level.

TABLE 6A

Results of Significance Tests from Logit Equations for the Choice of Natural Side in Penalty Shoot-outs

Panel A: Testing for Homogeneity of Goalkeepers

Dependent Variable:	Goal is scored	Kicker chooses Natural side	Goalkeeper goes to Natural side
P-values of *F*-Statistic for the joint insignificance of goalkeeper fixed-effects	0.57	0.71	0.69

Panel B: Logit Equation for the Choice of Natural Side in Goalkeepers

The estimating equation is $N = G[a_0 + a_1 lag(N) + a_2 lag2(N) + b_0 N^* + b_1 lag(N^*) + b_2 lag2(N^*) + c_1 lag(N) lag(N^*) + c_2 lag2(N) lag2(N^*)]$, where N and N^* denote the choice of "natural" strategy by a kicker and a goalkeeper, respectively (right for a right-footed kicker and for a goalkeeper facing a right-footed kicker, and left for a left-footed kicker and for a goalkeeper facing a left-footed kicker). The terms 'lag' and 'lag2' refer to the strategies previously followed in the ordered sequence of penalty kicks. G[x] denotes the function exp(x)/[(1+exp(x)]. Rejections are based on likelihood-ratio tests.

Null Hypothesis:	Likelihood-ratio test statistic (*p*-value)
1. $a_1 = a_2 = b_0 = b_1 = b_2 = c_1 = c_2 = 0$	6.17 (*p*-value: 0.40)
2. $a_1 = a_2 = 0$	2.03 (*p*-value: 0.36)
3. $b_1 = b_2 = 0$	1.71 (*p*-value: 0.42)
4. $c_1 = c_2 = 0$	1.03 (*p*-value: 0.59)
5. $b_0 = 0$	1.07 (*p*-value: 0.30)