

PENALTY KICKS IN SOCCER: AN ALTERNATIVE METHODOLOGY FOR TESTING MIXED-STRATEGY EQUILIBRIA

by Germán Coloma (CEMA University, Buenos Aires, Argentina)

Abstract

This paper uses the model proposed by Chiappori, Levitt and Groseclose (2002) to test for mixed-strategy Nash equilibria in a game between a kicker and a goalkeeper, in a situation of a soccer penalty kick. The main contribution of this paper is to develop an alternative testing methodology, based on the use of a simultaneous-equation regression approach that directly tests the equilibrium conditions. Using the same data than Chiappori, Levitt and Groseclose we find similar results than them, and we are also able to separately analyze the behavior of different types of players (classified according to the foot that they use to kick the ball, and to the strategies that they choose to mix).

JEL Classification: C72 (non-cooperative games), L83 (sports), C35 (simultaneous-equation discrete-regression models).

Keywords: soccer penalty kicks, mixed strategies, Nash equilibrium, simultaneous-equation regression, Wald test.

UNA METODOLOGÍA ALTERNATIVA PARA TESTEAR EQUILIBRIOS EN ESTRATEGIAS MIXTAS UTILIZANDO DATOS DE TIROS PENALES EN EL FÚTBOL

por Germán Coloma (Universidad del CEMA, Buenos Aires, Argentina)

Resumen

Este artículo utiliza el modelo propuesto por Chiappori, Levitt y Groseclose (2002) para testear equilibrios de Nash en estrategias mixtas en un juego entre un pateador y un arquero, en una situación de tiro penal en el fútbol. La mayor contribución del trabajo es que desarrolla una metodología alternativa, basada en el uso de un enfoque de regresión con ecuaciones simultáneas que testea directamente las condiciones de equilibrio del juego. Usando los mismos datos que Chiappori, Levitt y Groseclose se hallan resultados similares, pudiendo también analizarse separadamente el comportamiento de distintos tipos de jugadores (clasificados según el pie que usen para patear la pelota, y las estrategias que elijan mezclar).

Clasificación del JEL: C72 (juegos no cooperativos), L83 (deportes), C35 (modelos de regresión de ecuaciones simultáneas con variables discretas).

Descriptor: tiros penales en fútbol, estrategias mixtas, equilibrio de Nash, regresión de ecuaciones simultáneas, test de Wald.

PENALTY KICKS IN SOCCER: AN ALTERNATIVE METHODOLOGY FOR TESTING MIXED-STRATEGY EQUILIBRIA

by Germán Coloma*

1. Introduction

In a recent paper about mixed-strategy equilibrium testing, Chiappori, Levitt and Groseclose (2002) develop a model of a game between a goalkeeper and a kicker that can be applied to study the situation of a penalty kick in a soccer match. They also test that model using data from penalty kicks shot in the Italian and French first division leagues between 1997 and 2000. The aim of this paper is to continue that work, by developing an alternative and more direct methodology for testing mixed-strategy equilibria. Using the same database than Chiappori, Levitt and Groseclose, we apply that methodology and find some additional results that in general tend to confirm the original predictions. We are also able to classify the information according to different “types of players” and find that the results are consistent for most of the groups of data that we have.

2. The model

The model developed by Chiappori, Levitt and Groseclose describes the situation that the soccer players face in a penalty kick as a simultaneous-move game where the kicker has three alternative strategies: shooting right, left or center. Similarly, the goalkeeper also has three alternative strategies: moving to the right, moving to the left or remaining in the center of the goal. For reasons related to the clarity of the exposition (since the kicker’s right is the goalkeeper’s left, and viceversa) and to the fact that right-footed kickers and left-footed kickers generally have opposite tendencies when they decide where to shoot, we will use an alternative way to define the sides of the goal by referring to the “natural side” of the kicker (which is the goalkeeper’s right, if the kicker is right-footed, and the goalkeeper’s left, if the kicker is left-footed) and the “opposite side”. Labeled like that, the strategies of both kicker and goalkeeper will be to choose the natural side of the kicker (NS), the center (C) or the opposite side (OS).

Table 1: Scoring-probability matrix

		Goalkeeper		
		NS	C	OS
Kicker	NS	P_N	π_N	π_N
	C	μ	0	μ
	OS	π_O	π_O	P_O

The normal form of this game between a kicker and a goalkeeper can be represented through a scoring-probability matrix like the one that appears on table 1. This is because a penalty kick is a constant-sum game in which the objective of the kicker is to maximize that scoring probability, while the objective of the goalkeeper is to minimize it. Following Chiappori, Levitt and Groseclose’s model, we assume that, for a given penalty kick, the scoring probability can adopt six different values: π_N (which is the probability of scoring when the kicker chooses his natural side and the goalkeeper chooses the center or the opposite side), π_O (which is the probability of scoring when the kicker chooses the opposite side and the goalkeeper chooses the center or the natural side of the kicker), P_N (which is the probability of scoring when the kicker and the goalkeeper both choose the natural side of the kicker), P_O (which is the probability of scoring when the kicker and the goalkeeper both choose the opposite side), μ (which is the probability of scoring when the kicker chooses the

* CEMA University. Av. Córdoba 374, Buenos Aires (C1054AAP), Argentina. E-mail: gcoloma@cema.edu.ar. I thank Pierre-André Chiappori for having provided me the data that he used in his own study about this subject.

center and the goalkeeper chooses one of the sides), and zero (which is the probability of scoring when both players choose C). We also assume that “ $\pi_N > \pi_O > \mu > P_N > P_O > 0$ ”.

The game described in the previous paragraph has a unique Nash equilibrium, which always implies the use of mixed strategies by both players. Being a constant-sum strictly competitive game, this Nash equilibrium is also the minimax solution of the game (i.e., the situation in which each player is trying to minimize the maximal opposite player’s payoff). Depending on the relative values of the scoring probabilities, this equilibrium can be of two classes. One possible class is what Chiappori, Levitt and Grosseclose call “restricted randomization equilibrium” (RR), in which both the kicker and the goalkeeper choose NS and OS with positive probabilities, but they never choose C. The other class of equilibrium is what they call “general randomization equilibrium” (GR), in which both the kicker and the goalkeeper randomize over their three possible alternatives (NS, OS and C).

In an RR equilibrium, the probabilities that the kicker chooses NS (n) and OS (q) are respectively the following:

$$n = \frac{\pi_O - P_O}{\pi_N + \pi_O - P_N - P_O} \quad (1) ;$$

$$q = 1 - n = \frac{\pi_N - P_N}{\pi_N + \pi_O - P_N - P_O} \quad (2) ;$$

while the probabilities that the goalkeeper chooses NS (v) and OS (ω) are:

$$v = \frac{\pi_N - P_O}{\pi_N + \pi_O - P_N - P_O} \quad (3) ;$$

$$\omega = 1 - v = \frac{\pi_O - P_N}{\pi_N + \pi_O - P_N - P_O} \quad (4) .$$

On the other hand, if the equilibrium is GR, then the probabilities are:

$$n = \frac{\mu \cdot (\pi_O - P_O)}{(\pi_O - P_O) \cdot (\pi_N - P_N) + \mu \cdot (\pi_N + \pi_O - P_N - P_O)} \quad (5) ;$$

$$q = \frac{\mu \cdot (\pi_N - P_N)}{(\pi_O - P_O) \cdot (\pi_N - P_N) + \mu \cdot (\pi_N + \pi_O - P_N - P_O)} \quad (6) ;$$

$$c = 1 - n - q = \frac{(\pi_O - P_O) \cdot (\pi_N - P_N)}{(\pi_O - P_O) \cdot (\pi_N - P_N) + \mu \cdot (\pi_N + \pi_O - P_N - P_O)} \quad (7) ;$$

$$v = \frac{\pi_N \cdot (\pi_O - P_O) + \mu \cdot (\pi_N - \pi_O)}{(\pi_O - P_O) \cdot (\pi_N - P_N) + \mu \cdot (\pi_N + \pi_O - P_N - P_O)} \quad (8) ;$$

$$\omega = \frac{\pi_O \cdot (\pi_N - P_N) - \mu \cdot (\pi_N - \pi_O)}{(\pi_O - P_O) \cdot (\pi_N - P_N) + \mu \cdot (\pi_N + \pi_O - P_N - P_O)} \quad (9) ;$$

$$\gamma = 1 - v - \omega = \frac{\mu \cdot (\pi_N + \pi_O - P_N - P_O) + P_N \cdot P_O - \pi_N \cdot \pi_O}{(\pi_O - P_O) \cdot (\pi_N - P_N) + \mu \cdot (\pi_N + \pi_O - P_N - P_O)} \quad (10) ;$$

where “ c ” is the equilibrium probability that the kicker chooses C, and “ γ ” is the equilibrium probability that the goalkeeper chooses C.

One of the main theoretical results of Chiappori, Levitt and Grosseclose (2002) is that the unique Nash equilibrium of the game is RR if:

$$\mu \leq \frac{\pi_N \cdot \pi_O - P_N \cdot P_O}{\pi_N + \pi_O - P_N - P_O} \quad (11) ;$$

while it is GR if:

$$\mu > \frac{\pi_N \cdot \pi_O - P_N \cdot P_O}{\pi_N + \pi_O - P_N - P_O} \quad (12).$$

Both the RR and the GR equilibria have several properties that can be proven and tested. The most interesting are probably the ones that state that “ $n > q$ ”, “ $v > \omega$ ” and “ $v > n$ ”, and the one that states that, in a GR equilibrium, “ $c > \gamma$ ”. Like in any mixed-strategy equilibrium, in this one it also holds that the kicker’s and the goalkeeper’s randomization are independent, and that the equilibrium scoring probabilities are the same whether the kicker kicks NS or OS (or C, in a GR equilibrium) and whether the goalkeeper chooses NS or OS (or C, in a GR equilibrium).

3. The data

The data set used by Chiappori, Levitt and Groseclose consists of 459 penalty kicks, 242 of which were shot in games from the Italian first division league (between 1997 and 2000), and the other 217 were shot in games from the French first division league (between 1997 and 1999). For each penalty kick we know the name of the kicker, the name of the goalkeeper, their respective teams, which of those teams played home, if the kicker is either right-footed or left-footed, the score of the match at the time of the penalty kick, and the minute of the match at that time. We also know if the kicker shot NS, OS or C, and if the goalkeeper chose NS, OS or C, so we can classify the penalty kicks according to the players’ behavior into the nine cells of table 1 (NS-NS, NS-C, NS-OS, C-NS, C-C, C-OS, OS-NS, OS-C and OS-OS). We also know which penalty kicks were scored and which of them were not, so we can calculate a variety of scoring rates using different sorting criteria.

One possible classification to be applied to Chiappori, Levitt and Groseclose’s database is the one that divides the observations between kicks shot by right-footed kickers (384 observations) and kicks shot by left-footed kickers (75 observations)¹. Another possible classification has to do with observations generated by kickers and goalkeepers who choose restricted randomization strategies (i.e., players who never choose C as part of their mixed strategies) and observations generated by kickers and goalkeepers who choose general randomization strategies. Unlike the division based on the foot of the kicker, that classification cannot be directly observed, since for each kick we only know the actual realization of the strategy chosen (i.e., the place chosen by both the kicker and the goalkeeper) but not the full mixed strategy used. However, to divide our observations between RR and GR cases, we can use an indirect approximate approach, based on the cases in which we actually observe a kicker or a goalkeeper choosing the action C.

In Chiappori, Levitt and Groseclose’s database, there are 79 cases where kickers chose C, and 11 cases where goalkeepers chose C. Three of those cases overlap (i.e., they are cases where both the kicker and the goalkeeper chose C), so we have 87 observations where either the kicker or the goalkeeper have chosen the center of the goal². In order to classify the observations, we have considered all the penalty kicks shot by kickers included in that subset of 87 observations to be GR cases, and all the penalty kicks shot by kickers not included in that subset to be RR cases. This implies that 251 observations belong to the GR group, while the other 208 observations belong to the RR group.

The main statistics of the database used are summarized on table 2. By looking at the kicker’s and goalkeeper’s frequencies reported we can approximate the strategies actually played by the different types of players, using alternative modes of classification. This can

¹ This division between right-footed and left-footed kickers is slightly different from the one used by Chiappori, Levitt and Groseclose in their paper. The reason is that there is a penalty kick shot by a right-footed kicker (Palmieri, from the Italian team Lecce) which is reported to have been shot with the left foot. Chiappori, Levitt and Groseclose classify that shot as an observation generated by a left-footed kicker who shot to his natural side. Conversely, we are considering it as an observation generated by a right-footed kicker who shot to the opposite side.

² Notably, the three cases where both the kicker and the goalkeeper have chosen C are penalty kicks that were saved by the goalkeeper. This is consistent with the idea embedded in the model that the scoring probability in the cell C-C is zero.

serve as first approximations for the values of the parameters n , q , c , v , ω and γ , that appear in the theoretical model. The relationships among those parameters predicted by the model are in general confirmed by the frequencies reported, although there are a few details that are worth noting. For example, the Italian kickers shoot OS slightly more than NS, and the goalkeepers facing home team kickers seem to choose C much more frequently than the goalkeepers facing away team kickers.

Table 2: Basic statistics of the penalty-kicks database

Concept	# of obs	Kicker's frequency			Goalkeeper's frequency			Scoring rate
		NS	C	OS	NS	C	OS	
All penalty kicks	459	0.4466	0.1721	0.3813	0.5643	0.0240	0.4118	0.7495
Italy	242	0.4091	0.1653	0.4256	0.5413	0.0207	0.4380	0.7355
France	217	0.4885	0.1797	0.3318	0.5899	0.0276	0.3825	0.7650
Home team kickers	304	0.4276	0.1842	0.3882	0.5493	0.0132	0.4375	0.7500
Away team kickers	155	0.4839	0.1484	0.3677	0.5935	0.0452	0.3613	0.7484
Right-footed kickers	384	0.4375	0.1667	0.3958	0.5703	0.0234	0.4063	0.7656
Left-footed kickers	75	0.4933	0.2000	0.3067	0.5333	0.0267	0.4400	0.6667
RR cases	208	0.5433	0.0000	0.4567	0.6154	0.0000	0.3846	0.7644
GR cases	251	0.3665	0.3147	0.3187	0.5219	0.0438	0.4343	0.7371
Right-footed / RR	171	0.5205	0.0000	0.4795	0.6140	0.0000	0.3860	0.7661
Right-footed / GR	213	0.3709	0.3005	0.3286	0.5352	0.0423	0.4225	0.7653
Left-footed / RR	37	0.6486	0.0000	0.3514	0.6216	0.0000	0.3784	0.7568
Left-footed / GR	38	0.3421	0.3947	0.2632	0.4474	0.0526	0.5000	0.5789

Source: own elaboration based on Chiappori, Levitt and Groseclose (2002).

If we look at the scoring rates reported in the last column of table 2, we also see that in general the differences are small. The scoring rate is slightly higher in France than in Italy, and also slightly higher in the RR cases than in the GR cases. Both rates are almost equal when we compare penalty kicks shot by home and away teams, although the number of kicks conceded to home teams is notably higher than the number conceded to away teams (304 versus 155). There are also much more kicks shot by right-footed kickers than kicks shot by left-footed kickers, but that seems to be a natural consequence of the fact that the majority of the players are right-footed. However, what does not seem to be that natural is that the scoring rate of the left-footed kickers is considerably lower than the scoring rate of the right-footed kickers, and that difference is almost entirely explained by the group of left-footed kickers included in the GR type.

4. Estimation and testing procedures

Chiappori, Levitt and Groseclose (2002) apply several indirect procedures to test the existence of mixed-strategy equilibria in their penalty-kick data set. First of all, they run a regression to test the assumption that kickers and goalkeepers move simultaneously, trying to see if there is a positive correlation between the cases where kickers and goalkeepers choose the same side. They cannot reject the hypothesis that the correlation is zero, so they conclude that the simultaneous-move model can be a good representation of the outcome observed. They also test the assumption that players play mixed strategies, and conclude that the observed frequencies are consistent with that assumption, which is reinforced by the fact that they find no serial correlation among the observations of the different players.

Another hypothesis tested by Chiappori, Levitt and Groseclose is that goalkeepers are identical, in the sense that they all choose the same strategy and have approximately the same scoring rate against any given kicker. This hypothesis is important, because it allows to divide the observations in groups using kickers' characteristics only, while treating the goalkeepers as a homogeneous group. The procedure used by Chiappori, Levitt and Groseclose to test this consists of running regressions where the dependent variables are

dummies for whether the kick is scored, whether the kicker chooses NS, whether the kicker chooses C and whether the goalkeeper chooses NS, and the independent variables include kicker-fixed and goalkeeper-fixed effects. As they find that the variables that measure the goalkeeper-fixed effects are jointly insignificant from zero, they conclude that they cannot reject the assumption that goalkeepers are identical, which is an assumption that we will also use in our analysis.

The last test that Chiappori, Levitt and Groseclose perform has to do with the idea that the probability of scoring faced by each player is the same, whether they choose NS, OS or C. Although the probability values that they find for this hypothesis are relatively low, they obtain better results when they restrict themselves to kickers with eight or more kicks, and find only one kicker in that condition for whom that hypothesis is rejected at a 10% probability level. This makes them conclude that the concept of mixed-strategy equilibrium is a reasonable one to explain the behavior observed in their database, since one of the basic implications of that concept is that all players should be indifferent among playing the different actions that belong to their mixed strategies.

In another paper about equilibrium testing using data from soccer penalty kicks, Palacios-Huerta (2003) runs some additional tests that have to do with two features of the mixed-strategy equilibrium. His data set is different from the one used by Chiappori, Levitt and Groseclose (it contains 1417 observations from several European countries during the period 1995-2000), but its main statistics are roughly the same. What he finds is that scoring probabilities are statistically identical across strategies chosen by players, and that players generate serially independent sequences and therefore ignore any possible strategic link between plays. In a certain sense his results are more robust than Chiappori, Levitt and Groseclose's, but his analysis contains a major simplification. This is given by the fact that Palacios-Huerta pools the actions NS and C into a single group both for the kicker and the goalkeeper, and therefore uses a model in which each player chooses between two actions instead of three.

In none of the papers mentioned, however, the mixed-strategy equilibrium is tested directly, in the sense that their authors do not test the results stated by equations (4) to (10). A possible way to do that is to run a system of linear probability equations which, for an RR equilibrium, will have the following form:

$$KNS = c_0*LEFT + c_1*HOME + c_2*ITALY + c_3*(MINUTE-45) + c_4*SCORELAG + c_5 \quad (13) ;$$

$$GNS = c_{30}*LEFT + c_{31}*HOME + c_{32}*ITALY + c_{33}*(MINUTE-45) + c_{34}*SCORELAG + c_{35} \quad (14) ;$$

$$GOAL = c_{60}*LEFT + c_{61}*HOME + c_{62}*ITALY + c_{63}*(MINUTE-45) + c_{64}*SCORELAG + c_{65}*KNS*GNS + c_{66}*KOS*GOS + c_{67}*KNS*GOS + c_{68}*KOS*GNS \quad (15) ;$$

while, for a GR equilibrium, will be as follows:

$$KNS = c_0*LEFT + c_1*HOME + c_2*ITALY + c_3*(MINUTE-45) + c_4*SCORELAG + c_6 \quad (16) ;$$

$$KOS = c_{10}*LEFT + c_{11}*HOME + c_{12}*ITALY + c_{13}*(MINUTE-45) + c_{14}*SCORELAG + c_{16} \quad (17) ;$$

$$GNS = c_{30}*LEFT + c_{31}*HOME + c_{32}*ITALY + c_{33}*(MINUTE-45) + c_{34}*SCORELAG + c_{36} \quad (18) ;$$

$$GOS = c_{40}*LEFT + c_{41}*HOME + c_{42}*ITALY + c_{43}*(MINUTE-45) + c_{44}*SCORELAG + c_{46} \quad (19) ;$$

$$GOAL = c_{60}*LEFT + c_{61}*HOME + c_{62}*ITALY + c_{63}*(MINUTE-45) + c_{64}*SCORELAG + c_{69}*KNS*GNS + c_{70}*KOS*GOS + c_{71}*KC*(1-GC) + c_{72}*KNS*(1-GNS) + c_{73}*KOS*(1-GOS) \quad (20) .$$

The meaning of the variables included in equations (13) to (20) is the following: KNS, KOS, KC, GNS, GOS and GC are dummy variables for the different actions of the kicker and

the goalkeeper, LEFT is a dummy variable for the kicker being left-footed, HOME is a dummy variable for the kicker being part of the home team, ITALY is a dummy variable for the game being played in the Italian league, MINUTE indicates the minute of the game when the penalty kick was shot, SCORELAG is the score difference between the teams immediately before the penalty kick was shot, and GOAL is a dummy variable for whether the kick was scored or not.

In these systems, the equations whose dependent variables are KNS, KOS, GNS and GOS are equations that attempt to describe the strategies chosen by the different players, controlling for several observable differences that have to do with the situation where the penalty kick was shot (kicker's shooting foot, country league, being home or away, minute of the match and score of the match). The constants of those equations are therefore estimators of the mixed strategy chosen, so that, in an RR equilibrium, c5 is an estimation for n , and c35 is an estimation for v . Similarly, in a GR equilibrium, c6 is an estimation for n , c16 is an estimation for q , c36 is an estimation for v , and c46 is an estimation for ω^3 .

On the other side, equations (15) and (20) attempt to estimate the scoring probabilities of the different cells of the payoff matrix of the game played by the kicker and the goalkeeper. That is why, in an RR equilibrium, c65 is an estimation for P_N , c66 is an estimation for P_O , c67 is an estimation for π_N , and c68 is an estimation for π_O . Similarly, in a GR equilibrium, c69 is an estimation for P_N , c70 is an estimation for P_O , c71 is an estimation for μ , c72 is an estimation for π_N , and c73 is an estimation for π_O .

Given the equivalencies expressed in the two previous paragraphs, a possible way of testing for a mixed-strategy equilibrium consists of running a joint Wald test of the restrictions implied by equations (1), (3), (5), (6), (8) and (9). For an RR equilibrium this means:

$$c5 = \frac{c68 - c66}{c67 + c68 - c65 - c66} \quad (21) ;$$

$$c35 = \frac{c67 - c66}{c67 + c68 - c65 - c66} \quad (22) ;$$

while for a GR equilibrium the restrictions to be tested are:

$$c6 = \frac{c71 \cdot (c73 - c70)}{(c73 - c70) \cdot (c72 - c69) + c71 \cdot (c72 + c73 - c69 - c70)} \quad (23) ;$$

$$c16 = \frac{c71 \cdot (c72 - c69)}{(c73 - c70) \cdot (c72 - c69) + c71 \cdot (c72 + c73 - c69 - c70)} \quad (24) ;$$

$$c36 = \frac{c72 \cdot (c73 - c70) + c71 \cdot (c72 - c73)}{(c73 - c70) \cdot (c72 - c69) + c71 \cdot (c72 + c73 - c69 - c70)} \quad (25) ;$$

$$c46 = \frac{c73 \cdot (c72 - c69) - c71 \cdot (c72 - c73)}{(c73 - c70) \cdot (c72 - c69) + c71 \cdot (c72 + c73 - c69 - c70)} \quad (26) .$$

5. Estimation and testing results

In order to test the results of our model using the available data, we have first run a system of least-square linear-probability regressions using equations (16)-(20), implicitly assuming that all the observations were generated by a GR mixed-strategy Nash equilibrium. We have labeled that system with the name EQUILIBR1, and show its main results on table 3.

³ These equivalencies, of course, are true for a certain normalization of the game. Due to the way in which we have defined the independent control variables, they correspond to a hypothetical situation of a penalty kick shot by a right-footed kicker from the away team in the French league, in the last minute of the first half of a game in which the score is tied.

Table 3: Regression results (EQUILIBR1 model)

Concept	Coefficient	Std. error	t-statistic	p-value
KNS equation				
Constant (n)	0.523293	0.051152	10.23019	0.0000
R-squared	0.011967			
KOS equation				
Constant (q)	0.331125	0.049884	6.637888	0.0000
R-squared	0.015530			
GNS equation				
Constant (v)	0.624292	0.051012	12.23819	0.0000
R-squared	0.012249			
GOS equation				
Constant (ω)	0.329189	0.050557	6.511226	0.0000
R-squared	0.015136			
GOAL equation				
Parameter P _N	0.675918	0.049009	13.79174	0.0000
Parameter P _O	0.476091	0.058454	8.144762	0.0000
Parameter μ	0.883113	0.058542	15.08499	0.0000
Parameter π _N	0.983706	0.054843	17.93680	0.0000
Parameter π _O	0.936259	0.054080	17.31260	0.0000
R-squared	0.203442			

As an alternative specification for this model, we have built another system of equations which we named EQUILIBR2, which is a combination of the models stated in equations (13)-(15) and (16)-(20). Using the classification of our observations between RR and GR cases that appears on table 1, we have created the dummy variable RR, and we have run the following system of regressions:

$$\text{KNS} = c_0 \cdot \text{LEFT} + c_1 \cdot \text{HOME} + c_2 \cdot \text{ITALY} + c_3 \cdot (\text{MINUTE}-45) + c_4 \cdot \text{SCORELAG} + c_5 \cdot \text{RR} + c_6 \cdot (1-\text{RR}) \quad (27) ;$$

$$\text{KOS} = c_{10} \cdot \text{LEFT} + c_{11} \cdot \text{HOME} + c_{12} \cdot \text{ITALY} + c_{13} \cdot (\text{MINUTE}-45) + c_{14} \cdot \text{SCORELAG} + (1-c_5) \cdot \text{RR} + c_{16} \cdot (1-\text{RR}) \quad (28) ;$$

$$\text{GNS} = c_{30} \cdot \text{LEFT} + c_{31} \cdot \text{HOME} + c_{32} \cdot \text{ITALY} + c_{33} \cdot (\text{MINUTE}-45) + c_{34} \cdot \text{SCORELAG} + c_{35} \cdot \text{RR} + c_{36} \cdot (1-\text{RR}) \quad (29) ;$$

$$\text{GOS} = c_{40} \cdot \text{LEFT} + c_{41} \cdot \text{HOME} + c_{42} \cdot \text{ITALY} + c_{43} \cdot (\text{MINUTE}-45) + c_{44} \cdot \text{SCORELAG} + (1-c_{35}) \cdot \text{RR} + c_{46} \cdot (1-\text{RR}) \quad (30) ;$$

$$\begin{aligned} \text{GOAL} = & c_{60} \cdot \text{LEFT} + c_{61} \cdot \text{HOME} + c_{62} \cdot \text{ITALY} + c_{63} \cdot (\text{MINUTE}-45) \\ & + c_{64} \cdot \text{SCORELAG} + c_{65} \cdot \text{KNS} \cdot \text{GNS} \cdot \text{RR} + c_{66} \cdot \text{KOS} \cdot \text{GOS} \cdot \text{RR} \\ & + c_{67} \cdot \text{KNS} \cdot \text{GOS} \cdot \text{RR} + c_{68} \cdot \text{KOS} \cdot \text{GNS} \cdot \text{RR} + c_{69} \cdot \text{KNS} \cdot \text{GNS} \cdot (1-\text{RR}) \\ & + c_{70} \cdot \text{KOS} \cdot \text{GOS} \cdot (1-\text{RR}) + c_{71} \cdot \text{KC} \cdot (1-\text{GC}) \cdot (1-\text{RR}) \\ & + c_{72} \cdot \text{KNS} \cdot (1-\text{GNS}) \cdot (1-\text{RR}) + c_{73} \cdot \text{KOS} \cdot (1-\text{GOS}) \cdot (1-\text{RR}) \end{aligned} \quad (31) .$$

As we see, this system is useful to simultaneously test the RR and GR equilibria, using the assumption that 208 observations were generated by the first type of equilibrium and the other 251 observations were generated by the second type. This classification can be further refined by separating the right-footed from the left-footed kickers, treating them as separate types. That is what we have done in a third version of our model, labeled as EQUILIBR3. In that version we have eliminated the dummy variable LEFT from all the equations, but we have included combined dummy variables for right-footed RR cases, left-footed RR cases, right-footed GR cases and left-footed GR cases. We have also used those dummy variables in the scoring equation, in combination with the dummy variables designed

to estimate the corresponding scoring-probability coefficients. The estimated values for n , q , v , ω , P_N , P_O , μ , π_N and π_O that were obtained running all those alternative models are reported on table 4.

Table 4: Estimated parameters under different specifications

Concept	n	q	v	ω	P_N	P_O	μ	π_N	π_O
EQUILIBR1									
All kicks	0.52329	0.33113	0.62429	0.32919	0.67592	0.47609	0.88311	0.98371	0.93626
EQUILIBR2									
RR cases	0.59497	0.40503	0.67038	0.32962	0.64052	0.52923	---	0.98450	0.99611
GR cases	0.43012	0.23434	0.58136	0.36016	0.70670	0.40760	0.87395	0.96617	0.82353
EQUILIBR3									
Right-foot / RR	0.58280	0.41720	0.66159	0.33841	0.63672	0.48284	---	0.96313	0.98177
Left-foot / RR	0.73474	0.26526	0.67596	0.32404	0.58844	0.69597	---	1.00000	1.00000
Right-foot / GR	0.44873	0.22825	0.59102	0.35200	0.69788	0.39014	0.90939	0.97028	0.85751
Left-foot / GR	0.41417	0.17537 (*)	0.49690	0.43534	0.70622	0.43856 (**)	0.63014	0.86759	0.36158 (***)

Notes: (*) p-value: 0.0406; (**) p-value: 0.0179; (***) p-value: 0.1952; all the other coefficients reported have p-values smaller than 0.001.

To test for the mixed-strategy equilibria implied by equations (1) to (10) of the theoretical model, we have run Wald tests of the restrictions that appear on equations (21) to (26). Those tests were applied to the three alternative specifications of our model, and were both jointly and separately run for each kicker's type. Their results are reported on table 5.

Table 5: Mixed-strategy equilibrium tests

Concept	Chi-square	Probability
EQUILIBR1		
All penalty kicks	0.556377	0.967786
EQUILIBR2		
All penalty kicks	10.37535	0.109709
RR cases	1.520872	0.467463
GR cases	8.636000	0.070870
EQUILIBR3		
All penalty kicks	15.18509	0.231470
Right-footed / RR	0.757001	0.684888
Left-footed / RR	2.680897	0.261728
Right-footed / GR	6.974403	0.137247
Left-footed / GR	6.121931	0.190226

The results of the tests run are very favorable to the hypothesis that the data was generated by Nash mixed-strategy equilibria. This is notably the case for the most general model (EQUILIBR1), in which the null hypothesis cannot be rejected at any level of significance below 96%. The only case where the hypothesis can be rejected at a 10% level appears when we run the EQUILIBR2 model, and only for the GR cases. When we further refine our classification and run separate tests for the four types of kickers that we have identified, none of the tests can be rejected at a 10% level, and the global test for all the kickers has a probability value of 0.23147.

A last group of results that is worth report is the one that comes from running our three model specifications imposing the restrictions tested on the Wald tests. Those results are the ones that appear on table 6.

Table 6: Constrained estimated parameters under equilibrium hypotheses

Concept	n	q	v	ω	P_N	P_O	μ	π_N	π_O
EQUILIBR1									
All kicks	0.50786	0.32092	0.60937	0.31116	0.68635	0.48117	0.87008	0.97969	0.94538
EQUILIBR2									
RR cases	0.58773	0.41227	0.66681	0.33319	0.67744	0.47601	---	1.00000	0.94112
GR cases	0.48768	0.31268	0.58151	0.33283	0.61605	0.40490	0.82912	0.95547	0.93429
EQUILIBR3									
Right-foot / RR	0.57694	0.42306	0.65984	0.34016	0.65401	0.46018	---	1.00000	0.93247
Left-foot / RR	0.71006	0.28994	0.66787	0.33213	0.73541	0.50822	---	0.90972	0.93508
Right-foot / GR	0.49893	0.30379	0.58425	0.31889	0.63274	0.39470	0.85648	0.97139	0.95090
Left-foot / GR	0.49428	0.30360	0.52603	0.42838	0.49328	0.37217	0.64855	0.75848	0.80394

Note: all the coefficients reported are significantly different from zero at a 1% level.

The constrained estimated parameters reported on table 6 are remarkably good, in the sense that they satisfy almost all the theoretical assumptions of the model used. In five of the equilibria estimated it holds that " $\pi_N > \pi_O > \mu > P_N > P_O > 0$ ". The only two exceptions appear on the model EQUILIBR3, for which we got " $\pi_N < \pi_O$ " for the left-footed cases. In those cases, however, we cannot reject the hypothesis that " $\pi_N = \pi_O$ " for any reasonable level of significance, and we must also take into account that they refer to sub-samples that have only 37 and 38 observations.

The theoretical results of Chiappori, Levitt and Groseclose's model are also confirmed in almost all cases. The relationship " $\mu > (\pi_N \cdot \pi_O - P_N \cdot P_O) / (\pi_N + \pi_O - P_N - P_O)$ " holds for every GR equilibrium estimated, for which it also holds that " $c > \gamma$ " (i.e., the idea that kickers choose C more frequently than goalkeepers). In all the equilibria estimated, both RR and GR, we get that " $n > q$ " and " $v > \omega$ " (i.e., the idea that both kickers and goalkeepers choose NS more frequently than OS), and in all the equilibria except one (the left-footed RR case in the specification EQUILIBR3) we also get that " $v > n$ " (i.e., the idea that goalkeepers choose NS more frequently than kickers)⁴.

6. Concluding remarks

Since Nash (1951) created the idea of non-cooperative games, the concept of mixed-strategy equilibrium played a very important role in game theory. This is because in many game situations the only possible (or the most plausible) Nash equilibria are mixed-strategy equilibria. The empirical testing of this concept, however, has always been problematic, since in most situations it is impossible to actually observe players playing mixed strategies or to distinguish their behavior from other pure-strategy alternatives.

To test the concept of mixed-strategy equilibrium, the game-theory literature has traditionally relied on laboratory experiments, with results that are generally disappointing⁵. That is why, in recent years, the literature has turned to use data generated by actual sports contests, from which it is possible to gather repeated information from real play. The first example of this literature is Walker and Wooders (2001), who analyzed strategies chosen by tennis players when they had to decide where to serve, using data collected at Wimbledon (the most important tennis tournament in the world). Chiappori, Levitt and Groseclose (2002) and Palacios-Huerta (2003) are therefore examples that go in the same direction, using data from penalty kicks in professional soccer.

In the three papers mentioned in the previous paragraph, however, the testing

⁴ In the left-footed RR case, however, we cannot reject the hypothesis that " $n = v$ " for any reasonable level of significance.

⁵ Chiappori, Levitt and Groseclose (2002), for example, cite papers by Brown and Rosenthal (1990), Rapoport and Boebel (1992) and Ochs (1995), among others. All of them generally reject the hypothesis that players choose mixed-strategies that generate a Nash equilibrium.

procedures used were clearly indirect, since their main efforts are put on testing whether the players obtain the same winning probability in every pure strategy that they mix. In this paper we have used a different approach, based on the estimation of a simultaneous-equation model whose parameters are both the winning probabilities and the probabilities of playing the pure strategies, chosen by the players. That approach allows us to directly test the equilibrium equations of the theoretical specification of the model. Using the same database than Chiappori, Levitt and Groseclose, the results obtained are fully consistent with the ones gotten by those authors.

Another contribution of this paper is the way in which we have tried to solve the heterogeneity problem embedded in the data. This was done by classifying players according to one directly observable characteristic (the foot that they use to kick the ball) and by the type of mixed strategies that they are presumably playing.

References

- Brown, James and Rosenthal, Robert (1990). "Testing the Minimax Hypothesis: A Reexamination of O'Neill's Game Experiment"; *Econometrica*, vol 58, pp 1065-1081.
- Chiappori, Pierre-André; Levitt, Steven and Groseclose, Timothy (2002). "Testing Mixed-Strategy Equilibria when Players are Heterogeneous: The Case of Penalty Kicks in Soccer"; *American Economic Review*, vol 92, pp 1138-1151.
- Nash, John (1951). "Non-Cooperative Games"; *Annals of Mathematics*, vol 54, pp 286-295.
- Ochs, Jack (1995). "Games with Unique, Mixed Strategy Equilibria: An Experimental Study"; *Games and Economic Behavior*, vol 10, pp 202-217.
- Palacios-Huerta, Ignacio (2003). "Professionals Play Minimax"; *Review of Economic Studies*; vol 70, pp 395-415.
- Rapoport, Amnon and Boebel, Richard (1992). "Mixed Strategies in Strictly Competitive Games: A Further Test of the Minimax Hypothesis"; *Games and Economic Behavior*, vol 4, pp 261-283.
- Walker, Mark and Wooders, John (2001). "Minimax Play at Wimbledon"; *American Economic Review*, vol 91, pp 1521-1538.