

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

By P.-A. CHIAPPORI, S. LEVITT, AND T. GROSECLOSE*

The concept of mixed strategy is a fundamental component of game theory, and its normative importance is undisputed. However, its empirical relevance has sometimes been viewed with skepticism. The main concern over the practical usefulness of mixed strategies relates to the “indifference” property of a mixed-strategy equilibrium. In order to be willing to play a mixed strategy, an agent must be indifferent between each of the pure strategies that are played with positive probability in the mixed strategy, as well as any combination of those strategies. Given that the agent is indifferent across these many strategies, there is no benefit to selecting precisely the strategy that induces the opponent to be indifferent, as required for equilibrium. Why an agent would, in the absence of communication between players, choose exactly one particular randomization is not clear.¹

Of course, whether agents, in real life, *actually* play Nash equilibrium mixed strategies is ultimately an empirical question. The evidence to date on this issue is based almost exclusively on laboratory experiments (e.g., Barry O’Neill, 1987; Amnon Rapoport and Richard B. Boebel, 1992; Dilip Mookherjee and Barry Sopher,

1994; Jack Ochs, 1995; Kevin A. McCabe et al., 2000). The results of these experiments are mixed. O’Neill (1987) concludes that his experimental evidence is consistent with Nash mixed strategies, but that conclusion was contested by James N. Brown and Robert W. Rosenthal (1990). With the exception of McCabe et al. (2000), which looks at a three-person game, the other papers generally reject the Nash mixed-strategy equilibrium.

While much has been learned in the laboratory, there are inherent limitations to such studies. It is sometimes argued that behavior in the simplified, artificial setting of games played in such experiments need not mimic real-life behavior. In addition, even if individuals behave in ways that are inconsistent with optimizing behavior in the lab, market forces may discipline such behavior in the real world. Finally, interpretation of experiments rely on the assumption that the subjects are maximizing the monetary outcome of the game, whereas there may be other preferences at work among subjects (e.g., attempting to avoid looking foolish) that distort the results.²

Tests of mixed strategies in nonexperimental data are quite scarce. In real life, the games played are typically quite complex, with large strategy spaces that are not fully specified *ex ante*. In addition, preferences of the actors may not be perfectly known. We are aware of only one paper in a similar spirit to our own research. Using data from classic tennis matches, Mark Walker and John Wooders (2001) test whether the probability the player who serves the ball wins the point is equal for serves to the right and

* Chiappori: Department of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637; Levitt: Department of Economics, University of Chicago; Groseclose: Graduate School of Business, Stanford University, 518 Memorial Way, Stanford, CA 94305. The paper was presented at Games 2000 in Bilbao and at seminars in Paris and Chicago. We thank D. Braun, J. M. Conterno, R. Guesnerie, D. Heller, D. Mengual, P. Reny, B. Salanié, and especially J. L. Ertori for comments and suggestions, and M. Mazzocco and F. Bos for excellent research assistance. Any errors are ours.

¹ The theoretical arguments given in defense of the concept of mixed-strategy equilibria relate either to purification (John C. Harsanyi, 1973), or to the minimax property of the equilibrium strategy in zero-sum games. For recent elaborations on these ideas, see Authur J. Robson (1994) and Phil Reny and Robson (2001).

² The ultimatum game is one instance in which experimental play of subjects diverges substantially from the predicted Nash equilibrium. Robert Slonim and Alvin E. Roth (1998) demonstrate that raising the monetary payoffs to experiment participants induces behavior closer to that predicted by theory, although some disparity persists.

left portion of the service box, as would be predicted by theory. The results for tennis serves is consistent with equilibrium play.³

In this paper, we study penalty kicks in soccer. This application is a natural one for the study of mixed strategies. First, the structure of the game is that of “matching pennies,” thus there is a unique mixed-strategy equilibrium. Two players (a kicker and a goalie) participate in a zero-sum game with a well-identified strategy space (the kicker’s possible actions can be reasonably summarized as kicking to either the right, middle, or left side of the goal; the goalie can either jump to the right or left, or remain in the middle). Second, there is little ambiguity to the preferences of the participants: the kicker wants to maximize the probability of a score and the goalie wants to minimize scoring. Third, enormous amounts of money are at stake, both for the franchises and the individual participants. Fourth, data are readily available and are being continually generated. Finally, the participants know a great deal about the past history of behavior on the part of opponents, as this information is routinely tracked by soccer clubs.

We approach the question as follows. We begin by specifying a very general game in which each player can take one of three possible actions {left, middle, right}. We make mild general assumptions on the structure of the payoff (i.e., scoring probabilities) matrix; e.g., we suppose that scoring is more likely when the goalie chooses the wrong side, or that right-footed kickers are better when kicking to the left.⁴ The model is tractable, yet rich enough to generate complex and sometimes unexpected predictions. The empirical testing of these predictions raises very interesting aggregation problems. Strictly speaking, the payoff matrix is match-specific (i.e., varies depending on the identities of the goalie and the kicker). In our

data, however, we rarely observe multiple observations for a given pair of players.⁵ This raises a standard aggregation problem. While the theoretical predictions hold for *any* particular matrix, they may not be robust to aggregation; i.e., they may fail to hold *on average* for an heterogeneous population of games. We investigate this issue with some care. We show that several implications of the model are preserved by aggregation, hence can be directly taken to data. However, other basic predictions (e.g., equality of scoring probabilities across right, left, and center) do not survive aggregation in the presence of heterogeneity in the most general case. We then proceed to introduce additional assumptions into the model that provide a greater range of testable hypotheses. Again, these additional assumptions, motivated by the discussions with professional soccer players, are testable and cannot be rejected in the data.

The assumptions and predictions of the model are tested using a data set that includes virtually every penalty kick occurring in the French and Italian elite leagues over a period of three years—a total of 459 kicks. A critical assumption of the model is that the goalie and the kicker play simultaneously. We cannot reject this assumption empirically; the direction a goalie or kicker chooses on the current kick does not appear to influence the action played by the opponent. In contrast, the strategy chosen by a goalie today does depend on a kicker’s past history. Kickers, on the other hand, play as if all goalies are identical. We also find that all the theoretical predictions that are robust to aggregation (hence that can be tested directly on the total sample) are satisfied. Finally, using the result that goalies appear to be identical, we test, and do not reject, the null hypothesis that scoring probabilities are equal for kickers across right, left, and center. Also, subject to the limitations that aggregation imposes on testing goalie behavior, we cannot reject equal scoring probabilities with respect to goalies jumping right or left (goalies almost never stay in the

³ Much less relevant to our research is the strand of literature that builds and estimates game-theoretic models that sometimes involve simultaneous-move games with mixed-strategy equilibria such as Kenneth Hendricks and Robert Porter (1988) and Timothy F. Bresnahan and Peter C. Reiss (1990).

⁴ These general assumptions were suggested by common sense and by our discussions with professional soccer players. They are testable and supported by the data.

⁵ Even for a given match, the matrix of scoring probabilities may moreover be affected by the circumstances of the kick. We find, for instance, that scoring probabilities decline toward the end of the game.

middle). It is important to note, however, that some of our tests have relatively low power.

The remainder of the paper is structured as follows. Section I develops the basic model. Section II analyzes the complexities that arise in testing basic hypotheses in the presence of heterogeneity across kickers and goalies. We note which hypotheses are testable when the researcher has only a limited number of kicks per goalie–kicker pair, and we introduce and test restrictions on the model that lead to a richer set of testable hypotheses given the limitations of the data. Section III presents the empirical tests of the predictions of the model. Section IV concludes.

I. The Framework

A. Penalty Kicks in Soccer

According to the rule, “a penalty kick is awarded against a team which commits one of the ten offenses for which a direct free kick is awarded, inside its own penalty area and while the ball is in play.”⁶ The maximum speed the ball can reach exceeds 125 mph. At this speed, the ball enters the goal about two-tenths of a second after having been kicked. This means that a keeper who jumps *after* the ball has been kicked cannot possibly stop the shot (unless it is aimed at him). Thus the goalkeeper must choose the side of the jump before he knows exactly where the kick is aimed.⁷ It is generally believed that the kicker must also decide on the side of his kick before he can see the keeper move. A goal may be scored directly from a penalty kick, and it is actually scored in about

four kicks out of five.⁸ Given the amounts of money at stake, the value of any factor affecting the outcome even slightly is large.

In all first-league teams, goalkeepers are especially trained to save penalty kicks, and the goalie’s trainer keeps a record of the kicking habits of the other teams’ usual kickers. Conventional wisdom suggests that a right-footed kicker (about 85 percent of the population) will find it easier to kick to his left (his “natural side”) than his right; and vice versa for a left-footed kicker. The data strongly support this claim, as will be demonstrated. Thus, throughout the paper we focus on the distinction between the “natural” side (i.e., left for a right-footed player, right for a left-footed player) and the “nonnatural” one. We adopt this convention in the remainder. For the sake of readability, however, we use the terms “right” and “left” in the text, although technically these would be reversed for (the minority of) left-footed kickers.

B. The Model

Consider a large population of goalies and kickers. At each penalty kick, one goalie and one kicker are randomly matched. The kicker (respectively, the goalie) tries to maximize (minimize) the probability of scoring. The kicker may choose to kick to (his) right, his left, or the center of the goal. Similarly, the goalie may choose to jump to (the kicker’s) left, right, or to remain at the center. When the kicker and the goalie choose the same side S ($S = R, L$), the goal is scored with some probability P_S . If the kicker chooses S ($S = R, L$) while the goalie either chooses the wrong side or remains at the center, the goal is scored with probability $\pi_S > P_S$. Here, $1 - \pi_S$ can be interpreted as the probability that the kick goes out or hits the post or the bar; the inequality $\pi_S > P_S$ reflects the fact that when the goalie makes the correct choice, not only can the kick go out, but in addition it can be saved. Finally, a kick at the

⁶ The ball is placed on the penalty mark, located 11 m (12 yds) away from the midpoint between the goalposts. The defending goalkeeper remains on his goal line, facing the kicker, between the goalposts until the ball has been kicked. The players other than the kicker and the goalie are located outside the penalty area, at least 9.15 m (10 yds) from the penalty mark; they cannot interfere in the kick.

⁷ According to a former rule, the goalkeeper was not allowed to move before the ball was hit. This rule was never enforced; in practice, keepers always started to move before the kick. The rule was modified several years ago. According to the new rule, the keeper is not allowed to move *forward* before the ball is kicked, but he is free to move laterally.

⁸ The average number of goals scored per game slightly exceeds two on each side. About one-half of the games end up tied or with a one-goal difference in scores. In these cases, the outcome of a penalty kick has a direct impact on the final outcome.

center is scored with probability μ when the goalie jumps to one side, and is always saved if the goalie stays in the middle. Technically, the kicker and the goalie play a zero-sum game. Each strategy space is $\{R, C, L\}$; the payoff matrix is given by:

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

It should be stressed that, in full generality, this matrix is match-specific. The population is characterized by some distribution $d\phi(P_R, P_L, \pi_R, \pi_L, \mu)$ of the relevant parameters. We assume that the specific game matrix at stake is known by both players before the kick; this is a testable assumption, and we shall see it is not rejected by the data. Finally, we assume both players move simultaneously. Again, this assumption is testable and not rejected.

We now introduce three assumptions on scoring probabilities, that are satisfied by all matches. These assumptions were suggested to us by the professional goalkeepers we talked to, and seem to be unanimously accepted in the profession.

ASSUMPTION SC (“Sides and Center”):

$$(SC) \quad \pi_R > P_L \quad \pi_L > P_R$$

$$(SC') \quad \pi_R > \mu \quad \pi_L > \mu.$$

ASSUMPTION NS (“Natural Side”):

$$(NS) \quad \pi_L \geq \pi_R \quad P_L \geq P_R.$$

ASSUMPTION KS (“Kicker’s Side”):

$$(KS) \quad \pi_R - P_R \geq \pi_L - P_L.$$

Assumption (SC) states first that, if the kicker knew with certainty which direction the goalie would jump, he would choose to kick to the other direction [relation (SC)]. Also, if the goalie jumps to the kicker’s left (say), the scoring probability is higher for a kick to the right than to the center [relation (SC)]. The natural side (NS) assumption requires that the kicker kicks better on his natural side, whether the keeper guesses the side correctly or not. Finally,

TABLE 1—OBSERVED SCORING PROBABILITIES, BY FOOT AND SIDE

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

(KS) states that not only are kicks to the natural side less likely to go out, but they are also less easy to save.⁹

These assumptions are fully supported by the data, as it is clear from Table 1. The scoring probability when the goalie is mistaken varies between 89 percent and 95 percent (depending on the kicking foot and the side of the kick), whereas it ranges between 43 percent and 64 percent when the goalkeeper makes the correct choice, substantiating relation (SC). Also, the scoring probability is always higher on the kicker’s natural side (Assumption NS), and the difference is larger when the goalie makes the correct choice (Assumption KS). Regarding (SC’), our data indicate that the scoring probability, conditional on the goalie making the wrong choice, is 92 percent for a kick to one side versus 84 percent for a kick in the middle.¹⁰

C. Equilibrium: A First Characterization

The game belongs to the “matching penny” family. It has no pure-strategy equilibrium, but

⁹ If the goalie makes the wrong choice, the kicker scores unless the kick is out, which, for side X ($X = L, R$), happens with probability $1 - \pi_X$. If the goalie guesses the correct side, failing to score means either that the kick is out (which, because of independence, occurs again with probability $1 - \pi_X$), or that the kick is saved. Calling s_X the latter probability, one can see that

$$P_X = \pi_X - s_X$$

so that (KS’) is equivalent to

$$s_R \geq s_L.$$

¹⁰ These results should however be interpreted with caution, since aggregation problems may arise (see below).

it always admits a unique mixed-strategy equilibrium, as stated in our first proposition.

PROPOSITION 1: *There exist a unique mixed-strategy equilibrium. If*

$$(C_{\mu}) \quad \mu \leq \frac{\pi_L \pi_R - P_L P_R}{\pi_R + \pi_L - P_L - P_R}$$

then both players randomize over $\{L, R\}$ (“restricted randomization”). Otherwise both players randomize over $\{L, C, R\}$ (“general randomization”).

The proof relies on straightforward (although tedious) calculations. The interested reader is referred to Chiappori et al. (2000).

In a restricted randomization (RR) equilibrium, the kicker never chooses to kick at the center, and the goalie never remains in the center. An equilibrium of this type obtains when the probability μ of scoring when kicking at the center is small enough. The scoring probability is identical for both sides:

$$\begin{aligned} \Pr(\text{score} | S = L) &= \Pr(\text{score} | S = R) \\ &= \frac{\pi_L \pi_R - P_L P_R}{\pi_L + \pi_R - P_L - P_R} \end{aligned}$$

whereas a kick in the middle scores with strictly smaller probability μ .¹¹ In a generalized randomization (GR) equilibrium, on the other hand, both the goalie and the kicker choose right, left, or in the middle with positive probability, and the equilibrium scoring probabilities are equal.

Thus, kickers do not kick to the center unless the scoring is large enough, whereas they always kick to the sides with positive probability. With heterogeneous matches, this creates a selection bias, with the consequence that the *aggregate* scoring probability (i.e., proportion to kicks actually scored) should be larger for kicks to the center. We shall see that this pattern is actually observed in our data.

¹¹ Also, if $\pi_R = \pi_L$, the goalie and the kicker play the same mixed strategy.

D. Properties of the Equilibrium

We now present several properties of the equilibrium that will be crucial in defining our empirical tests.

PROPOSITION 2: *At the unique equilibrium of the game, the following properties hold true:*

1. *The kicker’s and the goalie’s randomization are independent.*
2. *The scoring probability is the same whether the kicker kicks right, left, or center whenever he does kick at the center with positive probability. Similarly, the scoring probability is the same whether the goalie jumps right, left, or center whenever he does remain at the center with positive probability.*
3. *Under Assumption (SC), the kicker is always more likely to choose C than the goalie.*
4. *Under Assumption (SC), the kicker always chooses his natural side less often than the goalie.*
5. *Under Assumptions (SC) and (NS), the keeper chooses the kicker’s natural side L more often than the opposite side R.*
6. *Under Assumptions (SC) and (KS), the kicker chooses his natural side L more often than the opposite side R.*
7. *Under Assumptions (SC), (NS), and (KS), the pattern (L, L) (i.e., the kicker chooses L and the goalie chooses L) is more likely than both (L, R) and (R, L), which in turn are both more likely than (R, R).*

Properties 1 and 2 are standard characterizations of a mixed-strategy equilibrium. Properties 3 and 4 are direct consequences of the form of the matrix and of Assumption (SC), and provide wonderful illustrations of the logic of mixed-strategy equilibria. For instance, the kicker’s probability of kicking to the center must make the *goalie* indifferent between jumping or staying (and vice versa for the goalie). Now, kicking at the center when the keeper stays is very damaging for the kicker (the scoring probability is zero), so it must be the case that at equilibrium this situation is very rare (the goalie should stay very rarely). Conversely, from the goalie’s perspective, kicks to the cen-

ter are not too bad, even if he jumps [they are actually better than kicks to the opposite side by (SC)], hence their equilibrium probability is larger.

Finally, the same type of reasoning applies the statements 5, 6, and 7. Assume the goalie randomizes between R and L with equal conditional probabilities. By Assumption (NS), the kicker would then be strictly better off kicking L, a violation of the indifference condition; hence at equilibrium the goalie should choose L more often. Similarly, Assumption (KS) implies that, should the kicker randomize equally between L and R, jumping to the right would be more effective from the goalie's viewpoint. Again, indifference requires more frequent kicks to the left. In all cases, the key remark is that the *kicker's* scoring probabilities are relevant for the *keeper's* strategy (and conversely), a conclusion that is typical of mixed-strategy equilibria, and sharply contrasts with standard intuition.

II. Heterogeneity and Aggregation

The previous propositions apply to any particular match. However, match-specific probabilities are not observable; only populationwide averages are. With a homogeneous population (i.e., assuming that the game matrix is identical across matches) this would not be a problem, since populationwide averages exactly reflect probabilities. Homogeneity, however, is a very restrictive assumption, that does not fit the data well (as demonstrated below). Heterogeneity will arise if players have varying abilities or characteristics, and may even be affected by the environment (time of the game, field condition, stress, fatigue, etc.). Then, a natural question is: which of the predictions above are preserved by aggregation, even in the presence of some arbitrary heterogeneity?

The following result summarizes the predictions of the model that are preserved by aggregation:

PROPOSITION 3: *For any distribution $d\Phi(P_R, P_L, \pi_R, \pi_L, \mu)$, the following hold true, under Assumption (SC):*

- (i) *The total number of kicks to the center is larger than the total number of kicks for which the goalie remains at the center.*
- (ii) *The total number of kicks to the kicker's left is smaller than the total number of jumps to the (kicker's) left.*
- (iii) *If Assumption (NS) is satisfied for all matches, then the number of jumps to the left is larger than the number of jumps to the right.*
- (iv) *If Assumption (KS) is satisfied for all matches, then the number of kicks to the left is larger than the number of kicks to the right.*
- (v) *If Assumptions (NS) and (KS) are satisfied for all matches, then the pattern (L, L) (i.e., the kicker chooses L and the goalie chooses L) is more frequent than both (L, R) and (R, L), which in turn are both more frequent than (R, R).*

Other results, however, may hold for each match but fail to be robust to aggregation. For instance, the prediction that the scoring probability should be the same on each side does not hold on aggregate, even when it works for each possible match. Assume, for instance, that there are two types of players, who differ in ability and equilibrium side, say, the best players shoot relatively more often to the left at equilibrium. Then a left kick is more likely to come from a stronger player and therefore has a higher chance of scoring. Econometrically, this is equivalent to stating that a *selection bias* arises whenever the side of the kick is correlated with the scoring probabilities; and theory asserts it must be, since it is endogenously determined by the probability matrix.

The heterogeneity problem may arise even when the same kicker and goalie are matched repeatedly, since scoring probabilities are affected by various exogenous variables.¹² Therefore, the equal scoring probability property should *not* be tested on raw data, but instead *conditional on observables*.¹³ However,

¹² For instance, we find that the scoring probability is larger for a penalty kick during the first 15 minutes of the game, and smaller for the last half hour.

¹³ We find, however, that while scoring probabilities do change over time during the game, the probabilities of

conditioning on covariates is not enough. While the total number of kicks available is fairly large, they mostly represent different pairings of kickers and goalies. For any given pairing, there are at most three kicks, and often one or two (or zero). Match-specific predictions are thus very difficult to test. Two solutions exist at this point. First, it is possible to test the predictions that are preserved by aggregation. Second, specific assumptions on the form of the distribution will allow testing of a greater number of predictions. Of course, it is critical that these assumptions be testable and not rejected by the data. In what follows, we use the following assumption:

ASSUMPTION IG (Identical Goalkeepers): *For any match between a kicker i and a goalie j , the parameters P_R , P_L , π_R , π_L , and μ do not depend on j .*

In other words, while kickers differ from each other, goalies are essentially identical. The game matrix is kicker-specific, but it does not depend on the goalkeeper; for a given kicker, each kicker–goalie pair faces the same matrix whatever the particular goalie involved.

Note, first, that this assumption can readily be tested; as we shall see, it is not rejected by the data. Also, Assumption IG, if it holds true, has various empirical consequences.

PROPOSITION 4: *Under Assumption IG, for any particular kicker i , the following hold true:*

- (i) *The kicker's strategy does not depend on the goalkeeper.*
- (ii) *The goalkeeper's strategy is identical for all goalkeepers.*
- (iii) *The scoring probability is the same whether the kicker kicks right or left, irrespective of the goalkeeper. If the kicker kicks at the center with positive probab-*

ity, the corresponding scoring probability is the same as when kicking at either side, irrespective of the goalkeeper.

- (iv) *The scoring probability is the same whether the goalkeeper jumps right or left, irrespective of the goalkeeper. If the kicker kicks at the center with positive probability, the corresponding scoring probability is the same as when kicking at either side, irrespective of the goalkeeper.*
- (v) *Conditional on not kicking at the center, the kicker always chooses his natural side less often than the goalie.*

From an empirical viewpoint, Assumption (IG) has a key consequence: all the theoretical results, including those that are not preserved by aggregation, can be tested kicker by kicker, using all kicks by the same kicker as independent draws of the same game.

III. Empirical Tests

We test the assumptions and predictions of the model in the previous sections using a data set of 459 penalty kicks. These kicks encompass virtually every penalty kick taken in the French first league over a two-year period and in the Italian first league over a three-year period. The data set was assembled by watching videotape of game highlight films. For each kick, we know the identities of the kicker and goalie, the action taken by both kicker and goalie (i.e., right, left, or center), which foot the kicker used for the shot, and information about the game situation such as the current score, minute of the game, and the home team. A total of 162 kickers and 88 goalies appear in the data. As a consequence of the relatively small number of observations in the data set, some of our estimates are imprecise, leading our tests to have relatively low power to discriminate between competing hypotheses. Because the power of some of the tests of the model increases with the number of observations per kicker, in some cases we limit the sample to either the 41 kickers with at least four shots (58 percent of the total observations) or the nine kickers with at least eight shots (22 percent of the total observations).

kicking to the right or to the left are not significantly affected. This suggests that the bias induced by aggregation over games with different covariates may not be too severe.

A. Testing the Assumption That Kickers and Goalies Move Simultaneously

Before examining the predictions of the model, we first test the fundamental assumption of the model: the kicker and goalie move simultaneously. Our proposed test of this assumption is as follows. If the two players move simultaneously, then conditional on the player's and the opponent's past history, the action chosen by the opponent on *this* penalty kick should not predict the other player's action on *this* penalty kick. Only if one player moves first (violating the assumption of a simultaneous-move game) should the other player be able to tailor his action to the opponent's actual choice on this particular kick. We implement this test in a linear probability regression of the following form:¹⁴

$$(SM) \quad R_i^K = \mathbf{X}_i \alpha + \beta R_i^G + \gamma \bar{R}_i^K + \delta \bar{R}_i^G + \varepsilon_i$$

where R_i^K (respectively, R_i^G) is a dummy for whether, in observation i , the kicker shoots (keeper jumps) right, \bar{R}_i^K (\bar{R}_i^G) is the proportion of kicks by the kicker (of jumps by the goalie) going right on all shots except this one,¹⁵ and \mathbf{X} is a vector of covariates that includes a set of controls for the particulars of the game situation at the time of the penalty kick: five indicators corresponding to the minute of the game in which the shot occurs, whether the kicker is on the home team, controls for the score of the game immediately prior to the penalty kick, and interaction terms that absorb any systematic differences in outcomes across leagues or across years within a league. The key parameter in this specification is β , the coefficient on whether the goalie jumps right on this kick. In a simultaneous move game, β should be equal to zero.

Results from the estimation of equation (SM) are presented in Table 2. The odd-numbered columns include all kickers; the even

columns include only kickers with at least four penalty kicks in the sample. Kickers with few kicks may not have well-developed reputations as to their choice of strategies.¹⁶ Columns 1 and 2 include only controls for the observed kicker and goalie behaviors. Columns 3 and 4 add in the full set of covariates related to the particulars of the game situation at the time of the penalty kick. The results in Table 3 are consistent with the assumption that the kicker and goalie move simultaneously. In none of the four columns can the null hypothesis that β equals zero be rejected. For the full sample of kickers with covariates included, the goalie jumps in the same direction that the shooter kicks 2.7 percent more frequently than would be expected. When only kickers with at least four penalty kicks in the sample are included the situation reverses, with goalies slightly more likely to jump in the wrong direction.¹⁷

A second observation that emerges from Table 2 is that strategies systematically differ across kickers: those kickers who more frequently kick right in the other observations in the data set are also more likely to kick right on this kick.¹⁸ On the other hand, there appears to be no relationship between the strategy that a kicker adopts today and the behavior of the goalie on other shots in the data. This latter finding is consistent with results we present later suggesting that kickers behave as if all goalies are identical.

¹⁶ In contrast to kickers, who may really have taken very few penalty kicks in their careers, all goalies have presumably participated in many prior penalty kicks. Although these penalty kicks are not part of our data set, presumably this more detailed past history is available to the clubs.

¹⁷ There is no particular reason for using the goalie's action as the left-hand-side variable and the kicker's action as a right-hand-side variable. In any case, virtually identical coefficients on β are obtained when the two variables are reversed.

¹⁸ Remember that in this and all other analyses in the paper we have reversed right and left for left-footed kickers to reflect the fact that there is a natural side that kickers prefer and that the natural side is reversed for left-footed kickers. The differences in strategies across kickers emerge much more strongly prior to the correction for left-footed kickers.

¹⁴ Probit regressions give similar results, although the interpretation of the coefficients is less straightforward.

¹⁵ Similar tests have been run using only penalty kicks prior to the one at stake. As in Table 2, we are unable to reject the null hypothesis of simultaneous moves.

TABLE 2—TESTING THE ASSUMPTION THAT THE KICKER AND GOALIE MOVE SIMULTANEOUSLY
(DEPENDENT VARIABLE: KICKER SHOOTS RIGHT)

Variable	(1)	(2)	(3)	(4)
Keeper jumps right	0.042 (0.052)	0.025 (0.062)	0.027 (0.052)	-0.026 (0.063)
Kicker's percentage of shots to the right, excluding this kick	0.219 (0.082)	0.370 (0.122)	0.220 (0.082)	0.357 (0.126)
Goalie's percentage of jumps to the right, excluding this kick	-0.032 (0.103)	0.001 (0.131)	-0.012 (0.104)	0.001 (0.135)
(League × year) dummies included?	yes	yes	yes	yes
Full set of covariates included?	no	no	yes	yes
Sample limited to kickers with 4+ kicks?	no	yes	no	yes
R^2 :	0.029	0.051	0.068	0.087
Number of observations:	373	252	373	252

Notes: The baseline sample includes all French first-league penalty kicks from 1997–1999 and all Italian first-league kicks (1997–2000) that involve a kicker and goalie each of whom have at least two kicks in the data set. If the kicker and goalie move simultaneously, then a goalie's action on this kick should not predict the kicker's action. At least two kicks are required so that the variables about goalie and kicker behavior on other penalty kicks can be constructed. Columns 2 and 4 limit the sample to kickers with at least four kicks in the sample. Regressions in columns 3 and 4 also include the following covariates not shown in the table: six indicator variables corresponding to 15-minute intervals of the game, whether the kicker is on the home team, and five indicators capturing the relative score in the game immediately prior to the penalty kick. None of the coefficients on these covariates is statistically significant at the 5-percent level. Standard errors are in parentheses. For shots involving left-footed kickers, the directions have been reversed so that shooting left corresponds to the “natural” side for all kickers.

TABLE 3—OBSERVED MATRIX OF SHOTS TAKEN

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

Notes: The sample includes all French first-league penalty kicks from 1997–1999 and all Italian first-league kicks (1997–2000). For shots involving left-footed kickers, the directions have been reversed so that shooting left corresponds to the “natural” side for all kickers.

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

Given that the kicker and goalie appear to move simultaneously, we shift our focus to testing the predictions of the model. We begin with those predictions of the model that are robust to aggregation across heterogeneous players.

Perhaps the most basic prediction of the

model is that all kickers and all goalies should play mixed strategies. Testing of this prediction is complicated by two factors. First, since we only observe a small number of plays for many of the kickers and goalies, it is possible that even if the player is employing a mixed strategy, only one of the actions randomized over will actually be observed in the data.¹⁹ On the other hand, if players use different strategies against different opponents, then multiple observations on a given player competing against different opponents may suggest that the player is using a mixed strategy, even if this is not truly the case. With those two caveats in mind, we first find that there are no kickers in our sample with at least four kicks who always kick in one direction. Only three of the 26 kickers with exactly three penalty kicks always shoot in the same direction. Even among kickers with exactly two shots, the same strategy is played both

¹⁹ The extreme case is when we have only one observation for a player, so that there is no information as to whether a mixed strategy is being used.

times in less than half the instances. Overall, there are 91 kickers in our sample with at least two kicks. Under the assumption that each of these kickers randomizes over the three possible strategies (left, middle, right) with the average frequencies observed in the data for all kickers, it is straightforward to compute the predicted number of kickers in our sample who should be observed always kicking the same direction, conditional on the number of kicks we have by kicker. We predict 14.0 (SE = 3.2) kickers should be observed playing only one strategy. In the actual data, this number is 16, well within one standard deviation of our predictions. Standard tests confirm that the observed frequencies match the theory quite well.

Results on goalies are essentially similar. The overwhelming majority of goalies with more than a few observations in the data play mixed strategies. There is, however, one goalie in the sample who jumps left on all eight kicks that he faces (only two of eight kicks against him go to the left, suggesting that his proclivity for jumping left is not lost on the kickers). Overall, we expect 9.9 (SE = 2.5) instances of observing only one strategy played, whereas there are 13 cases in the data.

Finally, an additional testable prediction of true randomizing behavior is that there should be no serial correlation in the strategy played. In other words, conditional on the overall probability of choosing left, right, or center, the actual strategy played on the previous penalty kick should not predict the strategy played this time. Consistent with this hypothesis, in regressions predicting the side that a kicker kicks or the goalie jumps in which we control for the average frequency with which a player chooses a side, the side played on the previous penalty kick by either the kicker or the goalie is never a statistically significant predictor of the side played on this shot by either player. This result is in stark contrast to past experimental studies (e.g., Brown and Rosenthal, 1990) and also to Walker and Wooders (2001) analysis of serves in tennis.²⁰

²⁰ The absence of serial correlation in our setting is perhaps not so surprising since the penalty kicks take place days or weeks apart. A more compelling test would involve

Table 3 presents the matrix of actions taken by kickers and goalies in the sample (the percentage of cases corresponding to each of the cells is shown in parentheses). There are five predictions of the model that can be tested using the information in Table 3. First, the model predicts that the kicker will choose to play “center” more frequently than the goalie (this is the content of Proposition 3(i) above). The result emerges very clearly in the data: kickers play “center” 79 times in the sample, compared to only 11 times for goalies.

A second prediction of the model is that goalies should play “left” (the kicker’s natural side) more frequently than kickers do. Indeed, goalies play “left” 200 times (56.6 percent of kicks), compared to 206 (44.9 percent) instances for kickers. Thus, the null that goalies play left more often than kickers cannot be rejected.²¹

The third and fourth predictions of the model are that under Assumptions (NS) and (KS), the kicker and the goalie are both more likely to go left than right. This prediction is confirmed: in the data, 260 jumps are made to the (kicker’s) left, and only 188 to the right. The same pattern holds for the kicker, although in a less spectacular way (206 against 174). Finally, given independence, a fifth prediction of the theory is that the cell “left-left” should have the greatest number of observations. This prediction is confirmed by the data, with the kicker and goalie both choosing left more than 25 percent of the time. The next most common outcome (goalie left, kicker right) appears about 20 percent of the time. Finally, the “right-right” pattern is the least frequent, as predicted by the model.

For completeness, Table 4 presents the matrix of scoring probabilities as a function of the actions taken by kickers and goalies. As noted

the choice of sides in World Cup tiebreakers, which involve consecutive penalty kicks for each side.

²¹ Actually, testing the null of equal propensities leads to rejection at the 10-percent level. This result is somewhat amplified by the fact that kickers play “middle” much more frequently than goalies. Even conditional on playing either “right” or “left,” goalies are more likely to choose “left” (58 percent for goalies versus 54 percent for kickers, although the difference is no longer significant). However, predictions about conditional probabilities are not robust to aggregation.

TABLE 4—OBSERVED MATRIX OF OUTCOMES:
PERCENTAGE OF SHOTS IN WHICH A GOAL IS SCORED

Goalie	Kicker			Total
	Left	Middle	Right	
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: The sample includes all French first-league penalty kicks from 1997–1999 and all Italian first-league kicks (1997–2000). For shots involving left-footed kickers, the directions have been reversed so that shooting left corresponds to the “natural” side for all kickers.

in the theory portion of this paper, with heterogeneous kickers or goalies, our model has no clear-cut predictions concerning the aggregate likelihoods of success. If kickers and goalies were all identical, however, then one would expect the average success rate for kickers should be the same across actions, and similarly for goalies. In practice the success probabilities across different actions are close, especially for goalies, where the fraction of goals scored varies only between 72.7 and 76.2 percent across the three actions. Interestingly, for kickers, playing middle has the highest average payoff, scoring over 80 percent of the time; this is exactly what was suggested by the “selection bias” argument developed above (see Section I, subsection C). Kicking right has the lowest payoff, averaging only 70 percent success.

C. Identical Goalkeepers

As demonstrated in the theory section of the paper, if goalies are identical, then we are able to generate additional predictions from our model. The assumption that goalies are homogeneous is tested in Table 5 using a regression framework. We examine four different outcome variables: the kick is successful, the kicker shoots right, the kicker shoots in the middle, and the goalie jumps right. Included as explanatory variables are the covariates describing the game characteristics used above, as well as goalie-fixed and kicker-fixed effects. The null hypothesis that all goalies are homogeneous corresponds to the goalie-fixed effects being

jointly insignificant from zero. In order to increase the power of this test, we restrict the sample to goalies with at least four penalty kicks in the data set. The F statistic for the joint test of the goalie-fixed effects is presented in the top row of Table 5. The cutoff values for rejecting the null hypothesis at the 10- and 5-percent level, respectively, are 1.31 and 1.42. In none of the four cases can we reject the hypothesis that all goalies are identical.²²

If goalies are indeed homogeneous, then a given kicker’s strategy will be independent of the goalie he is facing. This allows us to test the hypothesis that each kicker is indifferent across the set of actions that he plays with positive probability. We test this hypothesis by running linear probability models of the form

$$S_{i,t} = \mathbf{X}_{i,t}\boldsymbol{\alpha} + \sum \beta_i D_i + \sum \gamma_i D_i R_{i,t} + \sum \delta_i D_i M_{i,t} + \varepsilon_{i,t}$$

where $S_{i,t}$ is a dummy for whether the kick is scored, $S_{i,t}^*$ is the corresponding latent variable, D_i is a dummy for kicker i , $R_{i,t}$ (respectively, $M_{i,t}$) is a dummy for whether the kick goes right (middle), and \mathbf{X} is the same vector of covariates as before.

By including a fixed effect for each kicker, we allow each kicker to have a different probability of scoring. The test of the null hypothesis is that the vector of coefficients ($\gamma_1, \dots, \gamma_n, \delta_1, \dots, \delta_n$) are jointly insignificantly different from zero. The results of this test are presented in the top panel of Table 6. Results are shown separately for the set of kickers with five or more kicks in the sample (a total of 27 kickers) and the set of kickers with eight or more kicks in the sample (nine kickers). We report results with and without the full set of covariates included. If a player’s strategy is a function of observable characteristics such as the time of

²² In contrast, there is substantial evidence of heterogeneity across kickers. When we do not account for left-footed and right-footed kickers having their natural sides reversed, the homogeneity of all kickers is easily rejected. Once we make the natural foot adjustment, an F -test that all of the kicker-fixed effects are identical is rejected at the 10-percent level in two of the four columns in Table 5, when the sample is restricted to kickers with more than four kicks.

TABLE 5—TESTING WHETHER GOALIES ARE HOMOGENEOUS

Independent variable	Dependent variable			
	Kick successful	Kicker shoots right	Kicker shoots middle	Goalie jumps right
<i>F</i> statistic: joint significance of goalie-fixed effects [<i>p</i> value listed below]	0.95 [<i>p</i> = 0.57]	0.98 [<i>p</i> = 0.52]	0.88 [<i>p</i> = 0.70]	1.21 [<i>p</i> = 0.19]
Coefficients on other covariates:				
Minute 0–14	0.512 (0.134)	−0.220 (0.144)	0.113 (0.119)	0.134 (0.150)
Minute 15–29	0.291 (0.111)	0.049 (0.120)	0.043 (0.099)	0.047 (0.125)
Minute 30–44	0.254 (0.102)	0.038 (0.110)	0.083 (0.091)	0.030 (0.114)
Minute 45–59	0.124 (0.107)	0.082 (0.115)	0.105 (0.095)	0.026 (0.119)
Minute 60–74	0.105 (0.105)	0.014 (0.113)	0.098 (0.093)	0.003 (0.117)
(League × year) dummies included?	yes	yes	yes	yes
Kicker-fixed effects included?	yes	yes	yes	yes
Goalie-fixed effects included?	yes	yes	yes	yes
<i>R</i> ² :	0.552	0.571	0.532	0.557

Notes: The sample is limited to goalies with at least four penalty kicks in the data set. The first row presents an *F* test (with degrees of freedom equal to 50, 186) of the joint significance of the goalie-fixed effects. The *p* value of the *F* statistic is given in square brackets. If goalies are homogeneous, the *F* test should not reject the null hypothesis that all goalie-fixed effects are equal. All regressions also include controls for whether the kicker is on the home team and five indicators capturing the relative score in the game immediately prior to the penalty kick. None of the coefficients on the covariates that are not shown are statistically significant at the 5-percent level. Coefficient values and standard errors (in parentheses) are presented for other variables in the regression. The number of observations is 399. The omitted time category is the 75th minute of the game and beyond. For shots involving left-footed kickers, the directions have been reversed so that shooting left corresponds to the “natural” side for all kickers.

the game or the score of the game, then in principle these covariates should be included.²³ In none of the four columns can we reject the joint test of equality of scoring probabilities across strategies for kickers in the sample at the 5-percent level, although when covariates are not included the values are somewhat close to that cutoff. For individual kickers, we can reject equality across directions kicked at the 10-percent level in five of 27 cases in the sample of kickers with five or more kicks, whereas by chance one would expect only 2.7 values that extreme. Thus, there is evidence that a subset of individual kickers may not be playing optimally. In the sample restricted to kickers with eight or more kicks, only in one of nine cases is

an individual kicker beyond the 10-percent level, as would be expected by chance. While perhaps simply a statistical artifact, this result is consistent with the idea that those who more frequently take penalty kicks are most adapt at the randomization.

Given that kickers are not homogeneous, a direct test of goalie strategies along the lines presented in the top panel of Table 6 cannot be meaningfully interpreted. Under the maintained assumption that goalies are homogeneous, however, we can provide a different test. Namely, when facing a given kicker, goalies on average should in equilibrium obtain the same expected payoff regardless of which direction they jump. If all goalies are identical, then they should all play identical strategies when facing the same kicker. The bottom panel of Table 6 presents empirical evidence on the equality of scoring probabilities pooled across all goalies who face one of the kickers in our sample with at least eight kicks. The structure of the bottom panel of

²³ Note, however, that the manner in which we include the covariates is not fully general since we do not interact the covariates with the individual players; this is impossible because of the limited number of kicks per player in the sample.

TABLE 6—TESTING FOR IDENTICAL SCORING PROBABILITIES ACROSS LEFT, MIDDLE, AND RIGHT FOR INDIVIDUAL KICKERS AND THE GOALIES THEY FACE

Statistic	Kickers with five or more kicks		Kickers with eight or more kicks	
	(1)	(2)	(3)	(4)
A. Null hypothesis: <i>For a given kicker, the probability of scoring is the same when kicking right, middle, or left.</i>				
<i>P</i> value of joint test	0.10	0.28	0.15	0.45
<i>F</i> statistic	1.36	1.15	1.44	1.01
Degrees of freedom (numerator;denominator)	(43;136)	(43;123)	(16;76)	(16;63)
Number of individual kickers	27	27	9	9
Number of individual kickers for whom null is rejected at 0.10	5	5	1	1
Full set of covariates included in specification?	no	yes	no	yes
B. Null hypothesis: <i>For goalies facing a given kicker, the probability of scoring is the same whether the goalie jumps right or left</i>				
<i>P</i> value of joint test	0.31	0.28	0.42	0.19
<i>F</i> 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 of individual kickers for whom null is rejected at 0.10	5	4	1	1
Full set of covariates included in specification?	no	yes	no	yes

Notes: Statistics in the table are based on linear probability models in which the dependent variable is whether or not a goal is scored. The table assumes heterogeneity across kickers in success rates; that is, the hypothesis tested is whether, *for a given kicker*, success rates are identical when kicking right, middle, or left. No cross-kicker restrictions are imposed. The results in the bottom panel of the table refer to goalies facing a particular kicker, under the assumption that goalies are homogeneous. When included, the covariates are the same as those used elsewhere in the paper.

the table is identical to that of the top panel, except that the goalie's strategy replaces the kicker's strategy. The results are similar to that for kickers. In none of the four columns can the null hypothesis of equal probabilities of scoring across strategies be rejected for goalies at the 10-percent level.

IV. Conclusion

This paper develops a game-theoretic model of penalty kicks in soccer and tests the assumptions and predictions of the model using data from two European soccer leagues. The empirical results are consistent with the predictions of the model. We cannot reject that players optimally choose strategies, conditional on the opponent's behavior.

The application in this paper represents one of the first attempts to test mixed-strategy behavior using data generated outside of a controlled experiment. Although there are clear

advantages provided by a well-conducted laboratory experiment, testing game theory in the real world may provide unique insights. The penalty kick data we examine more closely corroborates the predictions of theory than past laboratory experiments would have led us to expect.

The importance of taking into account heterogeneity across actors plays a critical role in our analysis, since even some of the most seemingly straightforward predictions of the general model break down in the presence of heterogeneity. Carefully addressing the issue of heterogeneity will be a necessary ingredient of any future studies attempting to test game theory applications in real-world data.

REFERENCES

- Bresnahan, Timothy, F. and Reiss, Peter C. "Entry in Monopoly Markets." *Review of Economic Studies*, October 1990, 57(4), pp. 531–53.

- Brown, James N. and Rosenthal, Robert W.** "Testing the Minimax Hypothesis: A Reexamination of O'Neill's Game Experiment." *Econometrica*, September 1990, 58(5), pp. 1065–81.
- Chiappori, Pierre-Andre; Levitt, Steve and Groseclose, Timothy.** "Testing Mixed Strategy Equilibria When Players Are Heterogeneous: The Case of Penalty Kicks in Soccer." Working paper, University of Chicago, 2000.
- Harsanyi, John C.** "Games with Randomly Disturbed Payoffs: A New Rationale for Mixed-Strategy Equilibrium Points." *International Journal of Game Theory*, January 1973, 2(1), pp. 1–23.
- Hendricks, Kenneth and Porter, Robert.** "An Empirical Study of an Auction with Asymmetric Information." *American Economic Review*, December 1988, 78(5), pp. 865–83.
- McCabe, Kevin A.; Mukherji, Arijit and Runkle, David.** "An Experimental Study of Information and Mixed-Strategy Play in the Three-Person Matching-Pennies Game." *Economic Theory*, March 2000, 15(2), pp. 421–62.
- Mookherjee, Dilip and Sopher, Barry.** "Learning Behavior in an Experimental Matching Pennies Game." *Games and Economic Behavior*, July 1994, 7(1), pp. 62–91.
- Ochs, Jack.** "Games with Unique, Mixed Strategy Equilibria: An Experimental Study." *Games and Economic Behavior*, July 1995, 10(1), pp. 202–17.
- O'Neill, B.** "Nonmetric Test of the Minimax Theory of Two-Person Zerosum Games." *Proceedings of the National Academy of Sciences*, July 1987, 84(7), pp. 2106–09.
- Rapoport, Amnon and Boebel, Richard B.** "Mixed Strategies in Strictly Competitive Games: A Further Test of the Minimax Hypothesis." *Games and Economic Behavior*, April 1992, 4(2), pp. 261–83.
- Reny, Philip and Robson, Arthur J.** "Reinterpreting Mixed Strategy Equilibria: A Unification of the Classical and Bayesian Views." Working paper, University of Chicago, 2001.
- Robson, Arthur J.** "An 'Informationally Robust Equilibrium' in Two-Person Nonzero-Sum Games." *Games and Economic Behavior*, September 1994, 7(2), pp. 233–45.
- Slonim, Robert and Roth, Alvin E.** "Learning in High Stakes Ultimatum Games: An Experiment in the Slovak Republic." *Econometrica*, May 1998, 66(3), pp. 569–96.
- Walker, Mark and Wooders, John.** "Minimax Play at Wimbledon." *American Economic Review*, December 2001, 91(5), pp. 1521–38.