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

P.A. Chiappori† S. Levitt‡ T. Groseclose§

October 2000

*Paper presented at Games2000 in Bilbao and at seminars in Paris and Chicago. We thank J.M. Conterno, R. Guesnerie, D. Heller, D. Mengual, P. Reny, B. Salanié and especially J.L. Ettori for comments and suggestions, and M. Mazzocco and F. Bos for excellent research assistance. Errors are ours.

†Department of Economics, University of Chicago

‡Department of Economics, University of Chicago

§Graduate School of Business, Stanford University

Abstract

This paper tests the predictions of game theory using penalty kicks in soccer. Penalty kicks are modelled as a variant on "matching pennies" in which both the kicker and the goalie choose one of three strategies: left, middle, or right. We develop a general model allowing for heterogeneity across players and demonstrate that some of the most basic predictions of such a model do survive the aggregation necessary to test the model using real-world data, whereas others do not. We then present and test a set of assumptions sufficient to allow hypothesis testing using available data. The model yields a number of predictions, many of which are non-intuitive (e.g. that kickers choose middle more frequently than goalies). Almost all of these predictions are substantiated in data from the French and Italian soccer leagues. We cannot reject the null hypothesis that players are behaving optimally given the opponent's play.

1 Introduction: the positive content of mixed strategy equilibria

Mixed strategies are a fundamental component of game theory. As Nash (1950) demonstrates, every finite strategic form game has a (possibly degenerate) mixed strategy equilibrium. In games such as "matching pennies," no pure strategy equilibria exist.

While the normative importance of the concept of mixed strategy equilibrium is undisputed, its empirical relevance has sometimes been viewed with skepticism. There are two separate concerns relating to the practical usefulness of mixed strategies. The first criticism relates to the concept of Nash equilibrium in general. Since a mixed strategy is never a dominant strategy, whether a particular mixed strategy is the optimal choice against a given opponent depends critically on the strategy chosen by the opponent. This raises a coordination problem, the status of which is problematic. A second and more specific concern 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.

The theoretical arguments given in defense of the concept of mixed-strategy equilibria often relate to purification (Harsanyi 1973). Assume that each player's preferences are subject to small, random perturbations. In each occurrence of the game, each player knows her preferences but not the preferences of her opponent. For almost all realizations of the perturbations, the 'perturbed' game has a Nash Bayesian equilibrium in pure strategy. When the perturbations gradually vanish, any convergent sequence of pure strategy equilibria of the perturbed games converges to the mixed strategy equilibrium. Another justification is specific to zero-sum games, and stresses that in such games each player's equilibrium strategy coincide with the minimax solution. In other words, should a player adopt a cautious behavior and try to maximize the worst outcome she may get, then she will play her mixed strategy equilibrium; by so doing, she *strictly* maximizes her worst case payoff. Elaborating on these ideas, Robson (1994) introduces the concept of an 'informationally robust' equilibrium. Consider a two-person, zero-sum game where each agent, say i, believes that, with some probability ε, his opponent j will be able to *observe* i's strategy. Then one can again recover the mixed strategy equilibrium as a limit, with the additional property that, for each positive ε, the player's equilibrium mixed strategy is *strictly preferred* to any alternative[1].

Whether convincing or not, these argument are, at best, imperfect substitutes for empirical evidence. Ultimately, whether agents actually play mixed

[1] See Reny and Robson (2000) for an extension of these ideas to a more general class of games.

strategies corresponding to Nash equilibria is an empirical question. The evidence to date on this issue is based almost exclusively on laboratory experiments (e.g. O'Neill 1987, Rapoport and Boebel 1992, Ochs 1994, Mokherjee and Sopher 1997, 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 Brown and 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 non-experimental data are quite scarce. In real life, the games played are typically quite complex, with large strategy spaces

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

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, Walker and Wooders (Forthcoming) test whether the probability the player who serves the ball wins the point is equal for serves to the right and 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.

[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 Hendricks and Porter (1988) and Bresnahan and Reiss (1990).

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. Among the predictions of the model (besides randomization) are (a) the scoring probability is the same whether the kicker kicks right, left, or center, and similarly for the goalie's choice of strategy, (b) the kicker is more likely to choose center than the goalie, (c) the kicker will choose his "natural" side (the left side for right-footed kickers) less often than the goalie, (d) this is also true conditional on not kicking in the center, and (e) for right-footed kickers, the kicker and goalie both choosing left occurs with greater frequency than one choosing left and the other right, which in turn are more frequent than both players choosing right (and conversely for left-footed kickers).

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

[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

a typical aggregation problem. Although the predictions listed above 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 actually preserved by aggregation, hence can be directly taken to data[6]. However, other basic predictions - for instance, 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 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. As predicted by the model, kickers are more likely to choose center, less likely to choose their natural side than is the goalie, and (left, left) is the most common combination of actions whereas (right, right) is the least

the end of the game.

[6] These are predictions (b), (c) and (e) in the list above.

common. Finally, we cannot reject the null hypothesis that scoring probabilities are equal for kickers across right, left, and center. Subject to the limitations that aggregation imposes on testing goalie behavior, nor can we reject equal scoring probabilities with respect to goalies jumping right or left (goalies almost never stay in the 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 2 develops the basic model. Section 3 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 4 presents the empirical tests of the predictions of the model. Section 5 concludes.

2 The framework

2.1 Penalty kicks in soccer

According to the rule, 'a penalty kick is awarded against a team which commits one of the ten offences for which a direct free kick is awarded, inside its own penalty area and while the ball is in play'. 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]

The maximum speed the ball can reach exceeds 125 mph. At this speed, the ball enters the goal about two tenth 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 at[8]. Conversely, it is generally believed that the kicker has to decide on the side of his kick before he can see the keeper move. However, it has been argued that some kickers were able to wait till the last tenth of a second before deciding, and could thus sometimes be able to 'move second', i.e. to tailor their action to the goalie's actual choice - a crucial advantage in a matching pennies type of game. As we shall see, this feature does not seem to play a role in our data.

A goal may be scored directly from a penalty kick, and it is actually scored in about four kicks out of five. To understand the importance of a penalty kick, it must be remembered that 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 the kick has

[7] A precise statement of the rules can be found at http://www.drblank.com/slaw14.htm

[8] 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.

a direct impact on the final outcome. In addition, the evolution of the game is often affected by the penalty kick, resulting in an additional, indirect effect on the result. In almost two third of our sample, when the penalty kick takes place the kicker's team is either tied or lagging by one goal; then the outcome of the kick typically modifies the strategies followed by both teams. Finally, given the amounts of money at stake, the value of any factor affecting even slightly the outcome is large. Conservative estimates of the monetary consequences of an average penalty kick range from several thousands to several tens of thousands of dollars. But it can considerably exceed this amount on special occasions[9].

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. Interestingly enough, during training sessions a large percentage of kicks (more than 90%) are actually scored, whereas the proportion is smaller (around 80%) during games. This suggests that psychological aspects (or fatigue) may make a big difference in penalty kicks. Also, the physiology of the kicker plays an important role. While most kickers (almost 85%) are right-footed, some are left-footed. Of course, the respective role of right and left are inverted for a left-footed player. It is commonly believed, for instance, that a right-footed kicker will find it easier to kick at his left (his 'natural side') than his right; conversely, a left-footed player is expected to kick better at his

[9] In the 2000 final of the French cup, one team won 2-1 thanks to a penalty kick during the last five minuts of the game. Since the cup winner automatically qualifies for the European championship (UEFA cup) for the following year, the value of the kick probably exceeded ten millions dollars.

right. Considering only cases where the goalie jumps in the correct direction (by definition, a 'good' kick cannot be saved even when the goalie makes the correct choice), right-footed kickers score 65% of the time when shooting left, but only 42% of the time when going right. Left-footed kickers, on the other hand, exhibit the opposite pattern: the respective percentages are 54 % versus 57%. Kickers also do better going to their 'natural side' when the goalie chooses incorrectly.

These results should be taken with some care, because these percentages are computed from raw data without correction for selection bias - while theory states that, for each match, the choice of a side is endogenous to scoring probabilities. Still, they tend to confirm the general view of specialists.

As a consequence, the relevant distinction is between right and left, but between the 'natural' side (i.e., left for a right-footed player, right for a left-footed player) and the 'non-natural' one. Throughout the paper, we adopt this convention. Although, for the sake of readability, we keep the vocables of 'right' and 'left', the latter apply to right-footed players only, and should be switched for the (minority of) left-footed kickers.

2.2 The model

Consider a large population of goalies and kickers. At each penalty kick, one goalie and one kicker are randomly matched. The kicker (resp. the goalie) tries ot maximize (resp. minimize) the probability of scoring. The kicker may choose to kick to (his) right, his left, or to 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 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:

$_{K_i}\backslash^{G_j}$	L	C	R
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 $d\phi\left(P_R, P_L, \pi_R, \pi_L, \mu\right)$ 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 introduce two assumptions on scoring probabilities. These assumptions were suggested to us by the professional goalkeepers we talked to, and seem to be unanimously accepted in the profession. One is that scoring is always easier when the goalie makes the wrong choice. Natural as it seems, this assumption implies for instance that however good a kicker is at kicking to the left, he will still choose the right if he knows for sure that the goalie will go left. In the data, we find that the scoring probability when the goalie is mistaken varies between 80% and 95% (depending on the kicking foot and the side of the kick), whereas it ranges between 42% and 65% when the goalkeeper makes the correct choice. The second assumption is that conditional on the goalie choosing a particular side, the scoring probability is higher if the kick goes to the opposite side than if it is in the middle. In other words, a goalie jumping to his right (the kicker's left) is more likely to save a kick at the center than a kick to his left[10]. Again, this assumption is supported by the data: empirically, the scoring probability, conditional on the goalie making the wrong choice, is 92 % for a kick to one side versus 84 % for a kicks in the middle.

Formally:

Assumption *For any match, the parameters satisfy*

$$\pi_R \quad > \quad P_L \ , \ \pi_L > P_R \tag{A1}$$

and

$$\pi_R \quad > \quad \mu \ , \ \pi_L > \mu \tag{A2}$$

[10] The technical explanation is that while jumping to the right, a good goalkeeper can still save a kick in the middle, in particular with his left foot. A good kick at the center must thus be fast and high, which increases the probability of kicking on or over the bar.

2.3 Equilibrium: a first characterization

This game belongs to the 'matching penny' family, and has no pure strategy equilibrium. We thus introduce the following notations. For the kicker, l denotes the probability of choosing L and r the probability of choosing L; C is chosen with probability $1 - l - r$. The corresponding probabilities for the goalie are denoted using greek letters - i.e., λ, ρ and $1 - \lambda - \rho$. One can then readily characterize the equilibrium of the game:

Proposition 1 *There exist a unique mixed strategy equilibrium. If*

$$\mu \leq \frac{\pi_L \pi_R - P_L P_R}{\pi_R + \pi_L - P_L - P_R} \tag{C$_\mu$}$$

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

Proof : See Appendix

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. In that case, the mixed strategy of the kicker is defined by:

$$l_0 = \frac{\pi_R - P_R}{\pi_R + \pi_L - P_L - P_R} \tag{1}$$

$$r_0 = 1 - l_0 = \frac{\pi_L - P_L}{\pi_R + \pi_L - P_L - P_R} \tag{2}$$

whereas that of the goalie is

$$\lambda_0 = \frac{\pi_L - P_R}{\pi_L + \pi_R - P_L - P_R} \tag{3}$$

$$\rho_0 = 1 - \lambda_0 = \frac{\pi_R - P_L}{\pi_L + \pi_R - P_L - P_R} \tag{4}$$

In particular, conditional on side of the kick, the scoring probability is

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

whereas a kick in the middle scores with probability μ. If condition (C_μ) holds, the kicker will not try a center kick. Also, note that if $\pi_R = \pi_L$ then the goalie and the kicker play the same mixed strategy.

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 probabilities. Specifically, on the kicker's side, one gets

$$r = \frac{\mu(\pi_L - P_L)}{(\pi_R - P_R)(\pi_L - P_L) + \mu(\pi_R + \pi_L - P_R - P_L)} \tag{5}$$

$$l = \frac{\mu(\pi_R - P_R)}{(\pi_R - P_R)(\pi_L - P_L) + \mu(\pi_R + \pi_L - P_R - P_L)} \tag{6}$$

while for the goalie

$$\rho = \frac{\pi_R(\pi_L - P_L) - \mu(\pi_L - \pi_R)}{(\pi_R - P_R)(\pi_L - P_L) + \mu(\pi_R + \pi_L - P_R - P_L)} \tag{7}$$

$$\lambda = \frac{\pi_L(\pi_R - P_R) + \mu(\pi_L - \pi_R)}{(\pi_R - P_R)(\pi_L - P_L) + \mu(\pi_R + \pi_L - P_R - P_L)} \tag{8}$$

In particular, the goalie stays at the center with probability

$$1 - (\rho + \lambda) = \frac{\mu(\pi_R + \pi_L - P_R - P_L) - (\pi_R \pi_L - P_R P_L)}{(\pi_R - P_R)(\pi_L - P_L) + \mu(\pi_R + \pi_L - P_R - P_L)}$$

which is positive when condition (C_μ) holds.

The intuition behind these results is straightforward. Take a kicker who is extremely good at kicking to the left - say, P_L is close to π_L, meaning that the left kicks are so good that the goalie cannot do much even if he chooses the correct side. Such a kicker probably kicks more often to the left. However, *always* choosing left cannot be an equilibrium strategy, because the goalie would anticipate this pattern and always jump to the (kicker's) left. But then, 'deviating' by kicking to the right (or center) has an obvious advantage: the goalie won't be there, and the scoring probability is π_R (or μ). Since, by assumption, $\pi_R > P_L$, such a deviation is always profitable. That's why the equilibrium always entail randomization. Then the goalie will also jump right with some probability; which decreases the benefit of kicking to the right (the goalie might be there) and increase that of kicking to the left (the goalie might not be there). Equilibrium obtains when scoring probabilities are exactly equalized.

2.4 Properties of the equilibrium

These equations allow us to derive a few properties of the equilibrium probabilities. A first and very natural one states that the kicker's and keeper's strategies are independent, and that the scoring probabilities on each side should be equal:

Proposition 2 *At equilibrium, the kicker's and the goalie's randomization are independent. The scoring probability is the same whether the kicker kicks right, left, or center (whenever he does kick at the center). Similarly, the scoring probability is the same whether the goalie jumps right, left, or center (whenever* *he does remain at the center).*

In other words, the probability of observing any particular pattern - say, a kick to the right while the goalie jumps to the left - equals the product of the marginal probabilities - i.e., the probability that the kick goes to the right times the probability that the goalie jumps to the left. Also, goals are scored with equal likelihood for each pattern (provided that it occurs with positive probability). This last result is certainly expected, since this equality stems from the very definition of a mixed strategy equilibrium. Remember, however, that these prediction only hold for a particular game matrix, i.e. conditional on the realization of $(P_R, P_L, \pi_R, \pi_L, \mu)$.

A second set of results characterize the respective probabilities of choosing a particular move. We first derive the following result:

Proposition 3 *The kicker is always more likely to choose C than the goalie.*

Proof. Just note that

$$\begin{aligned} & [1-(r+l)]-[1-(\rho+\lambda)] \\ = & \frac{(\pi_R-P_R)(\pi_L-P_L)-\mu(\pi_R+\pi_L-P_R-P_L)+(\pi_R\pi_L-P_RP_L)}{(\pi_R-P_R)(\pi_L-P_L)+\mu(\pi_R+\pi_L-P_R-P_L)} \\ = & \frac{(\pi_L-\mu)(\pi_R-P_R)+(\pi_R-\mu)(\pi_L-P_L)}{(\pi_R-P_R)(\pi_L-P_L)+\mu(\pi_R+\pi_L-P_R-P_L)}>0 \end{aligned}$$

■

Intuitively, (A2) creates an asymmetry in the mistakes a goalie can make. Staying at the center is very costly when the kick goes to one side or the other, since the scoring probability is high. Conversely, jumping while the kick is at the center is less risky, since the scoring probability is smaller.

A second prediction concerns the probability of choosing L or R conditional on not kicking in the middle. Starting with the RR case, one has that

$$r_0 - \rho_0 = \frac{\pi_L - \pi_R}{\pi_R + \pi_L - P_L - P_R}$$

Considering now the GR case, from the formulas above:

$$\frac{r}{r+l} = \frac{\pi_L - P_L}{\pi_R + \pi_L - P_R - P_L}$$

while

$$\begin{aligned} \frac{\rho}{\lambda+\rho} &= \frac{\pi_R(\pi_L - P_L) + \mu(\pi_R - \pi_L)}{\pi_R(\pi_L - P_L) + \pi_L(\pi_R - P_R)} \\ &= \frac{\pi_L - P_L}{\pi_L - P_L + \frac{\pi_L}{\pi_R}(\pi_R - P_R)} - \frac{\mu(\pi_L - \pi_R)}{\pi_L - P_L + \frac{\pi_L}{\pi_R}(\pi_R - P_R)} \end{aligned}$$

It follows that

$$\rho_0 < r_0$$

and

$$\frac{\rho}{\lambda+\rho} < \frac{r}{r+l}$$

if and only if $\pi_R < \pi_L$.

We call a side S the kicker's 'natural side' if

$$\pi_S > \pi_{S'} \text{ for } S \neq S', \text{ where } S, S' \in \{R, L\}$$

We can state the following result

Proposition 4 *Conditional on not kicking at the center, the kicker always chooses his natural side less often than the goalie*

Another consequence of the previous result is that the kicker chooses the sides (as opposed to the center) less often than the goalie, and when he does he is less likely than the goalie to choose his natural side. Altogether, the number of kicks to the kicker's natural side must thus be smaller than the number of jumps to the right. Indeed, direct calculations show that (taking left as the natural side):

$$l - \lambda = \frac{-(\pi_L - \mu)(\pi_R - P_R) - \mu(\pi_L - \pi_R)}{(\pi_R - P_R)(\pi_L - P_L) + \mu(\pi_R + \pi_L - P_R - P_L)} < 0$$

Formally:

Corollary 5 *Unconditionally, the kicker always chooses his natural side less often than the goalie*

Concerning kicks to the non-natural side (here the right), as compared to jumps in the same direction, the answer depends on the type of randomization. In the RR case the kicker chooses R more often than the goalie. For GR, the situation is less clear, because two effects go in opposite directions: the kicker is more likely to choose R instead of L than the goalie, but he is also less likely to choose one of them (instead of C). Indeed:

$$r - \rho = \frac{-(\pi_R - \mu)(\pi_L - P_L) + \mu(\pi_L - \pi_R)}{(\pi_R - P_R)(\pi_L - P_L) + \mu(\pi_R + \pi_L - P_R - P_L)}$$

If the 'natural side' effect is weak (so that $\pi_L - \pi_R$ is small), the total number of R kicks is smaller than the total number of R jumps.

Finally, which side is actually chosen more often by each player is a very interesting question. Let us first consider the keeper. The answer is provided by the following proposition:

Proposition 6 *The keeper chooses the kicker's natural side S more often that the opposite side S' if and only if:*

$$\pi_S - \pi_{S'} + P_S - P_{S'} \geq 0 \tag{9}$$

Proof : In the RR case, the result is immediate. Consider, now, the GR case. Then

$$\lambda - \rho = \frac{-\pi_L P_R - 2\mu\pi_R + 2\mu\pi_L + \pi_R P_L}{(\pi_R - P_R)(\pi_L - P_L) + \mu(\pi_R + \pi_L - P_R - P_L)} = \frac{N}{D}$$

where

$$N = 2\mu(\pi_L - \pi_R) + \pi_R P_L - \pi_L P_R$$

and

$$D > 0$$

Assume, for instance, that L is the natural side, so that $\pi_L > \pi_R$. Since

$$\mu \geq \frac{\pi_L \pi_R - P_L P_R}{\pi_R + \pi_L - P_L - P_R}$$

one gets that

$$N \geq \frac{(\pi_L - \pi_R + P_L - P_R)(\pi_L(\pi_R - P_R) + \pi_R(\pi_L - P_L))}{\pi_R + \pi_L - P_R - P_L}$$

hence the conclusion. By symmetry, the same conclusion obtains, vice versa, if R is the natural side.

The result obtains in particular under a very simple assumption, namely that the kicker kicks better on his natural side irrespective of the side chosen by the keeper. Formally, this means that (assuming for instance that L is the natural side):

$$\pi_L \geq \pi_R \text{ and } P_L \geq P_R \tag{NS}$$

Throughout the paper, we refer to (NS) as the 'natural side' assumption. One can thus state:

Corollary 7 *If the natural side assumption (NS) holds, then the keeper jumps to the L more often that to the R*

During our discussions with professional soccer players, they all confirmed that the 'natural side' assumption had a very general relevance. Also, we shall see below that it is strongly supported by the data. One can thus confidently predict that the goalie will jump more often to the kicker's natural side than to the opposite side. This conclusion is particularly interesting in view of the purpose of the paper. The idea that the *kicker*'s scoring probabilities are more relevant for the *keeper*'s strategy is typical of mixed strategy equilibria, and sharply contrasts with standard intuition.

The case of the kicker is slightly more complex, as can be seen from the following result:

Proposition 8 *The kicker chooses L more often that R if and only if*

$$\pi_R - P_R \geq \pi_L - P_L \tag{KS}$$

Proof : Immediate from the previous equations

The latter result looks somewhat paradoxical, in that what matter for the kicker's choice is not the scoring probabilities per se, but rather the difference in scoring probabilities according to whether the keeper chooses the correct side or not. Remember, however, that the kicker's equilibrium strategy must be such that the *goalie* is indifferent between jumping R or L. What matters for the goalie is the 'cost of a mistake', i.e. the loss in saving probability resulting from a mistaken choice. That is exactly what equation (10) expresses.

The necessary condition (KS) can be rewritten as

$$P_L - P_R \geq \pi_L - \pi_R \tag{KS'}$$

(KS') states that the difference, in terms of scoring probability, between the kicker's natural side and the opposite side is particularly important when the goalie chooses the correct side. This assumption sounds quite sensible for the following reason. 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 $1 - \pi_X$. Assume, now, that 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 in but is saved. Calling s_X the latter probability, one can see that

$$P_X = \pi_X - s_X$$

so that (KS') is equivalent to

$$s_R \geq s_L$$

In words, not only kicks at the natural side are less likely to go out, but they are also less easy to save. Again, this assumption conforms to the professionals' opinion and is supported by the data.

One can thus summarize the previous discussion as follows:

Corollary 9 *Assume that both (NS) and (KS) hold true. Then 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)*.*

Finally, what is the empirical support of (NS) and (KS)? An answer is provided by Table 1, which that summarize the various observed scoring probabilities by foot and side.

Table 1: observed scoring probabilities, by foot and side

1.1 Right-footed kickers

Kicker \ Goalie	Correct side	Middle or wrong side
Left	64.9 %	94.6 %
Right	41.9 %	91.0 %

1.2 Left-footed kickers

Kicker \ Goalie	Correct side	Middle or wrong side
Left	53.8%	80.0%
Right	56.5%	93.3%

1.3 Total sample

Kicker \ Goalie	Correct side	Middle or wrong side
Natural side ('left')	63.6%	94.4%
Opposite side ('Right')	43.7%	89.3%

It is clear, from Table 1, that (NS) is strongly supported by the data for both right- and left-footed players. (KS) also holds on the total population: a kick to the natural side (as opposed to the opposite side) increases the scoring probability by 20 percentage points when the goalie makes the correct choice, against 5 when the goalie gets it wrong. Note that for left-footed kickers, the difference has the wrong sign but is not significant.[11]

3 Heterogeneity and aggregation

The previous propositions apply to any particular match. However, match-specific probabilities are not observable; only population-wide averages are. With an homogeneous population - i.e., assuming that the game matrix is identical across matches - this is not a problem, since population-wide averages exactly reflect probabilities. Then the model generates very strong predictions. Essentially, the basic parameters of the game (i.e., P_R, P_L, π_R, π_L and μ) can be directly estimated from scoring probabilities. One can then compute the predicted equilibrium strategies, and compare them to observed behavior.

Homogeneity, however, is a very restrictive assumption, that does not fit the data well (as it will be clear below). In principle, one should assume that the distribution of parameters is non degenerate. Then a natural question is: which of the predictions above are preserved by aggregation, irrespective of the form of the parameter distribution?

The answer depends on the particular property at stake. For instance,

[11] Altogether, about 15% of kickers are left-footed, and 30% of their kicks are aimed at their non-natural side; hence each second row cell of Table 1.2 represents less than 3% of the kicks (i.e., less than 15 kicks)

Proposition 3 holds true on aggregate, whatever the parameter distribution. Indeed, the total number of kicks at the center is

$$K_C = \int_{GR} \left[1 - \left(r\left(P_R, P_L, \pi_R, \pi_L, \mu\right) + l\left(P_R, P_L, \pi_R, \pi_L, \mu\right)\right)\right]\ d\phi\left(P_R, P_L, \pi_R, \pi_L, \mu\right)$$

where the set GR is defined by

$$GR = \left\{\left(P_R, P_L, \pi_R, \pi_L, \mu\right)\ \mid\ \mu \geq \frac{\pi_L \pi_R - P_L P_R}{\pi_L + \pi_R - P_L - P_R}\right\}$$

and where $r\left(P_R, P_L, \pi_R, \pi_L, \mu\right)$ and $l\left(P_R, P_L, \pi_R, \pi_L, \mu\right)$ are defined by (5) and (6). Similarly, the total number of kicks where the goalie stays at the center is

$$J_C = \int_{GR} \left[1 - \left(\rho\left(P_R, P_L, \pi_R, \pi_L, \mu\right) + \lambda\left(P_R, P_L, \pi_R, \pi_L, \mu\right)\right)\right]\ d\phi\left(P_R, P_L, \pi_R, \pi_L, \mu\right)$$

Hence

$$K_C - J_C = \int_{GR} \left[(1 - r - l) - (1 - \rho - \lambda)\right]\ d\phi \geq 0$$

since the integrand is positive over GR.

The same conclusion applies to Proposition 8 and Corollaries 5, 7 and 9. For instance, regarding Corollary 5:

$$K_L - J_L = \int_{RR} \left(l_0 - \lambda_0\right)\ d\phi + \int_{GR} (l - \lambda)\ d\phi < 0$$

where K_X (resp. J_X) denotes the number of kicks (resp. jumps) on side X, and where the set RR is defined by

$$RR = \left\{\left(P_R, P_L, \pi_R, \pi_L, \mu\right)\ \mid\ \mu < \frac{\pi_L \pi_R - P_L P_R}{\pi_L + \pi_R - P_L - P_R}\right\}$$

In the same way, assuming that assumptions (NS) and (KS) hold for all matches, one gets, for instance, that

$$J_{LL} - J_{RL} = \int_{RR} (l_0 - r_0)\, \lambda_0 \, d\phi + \int_{GR} (l - r)\, \lambda \, d\phi \geq 0$$

since the integrand is always positive.

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

Proposition 10 *For any distribution* $d\phi\,(P_R, P_L, \pi_R, \pi_L, \mu)$*:*

- *the total number of kicks to the center is larger than the total number of kicks for which the goalie remains at the center*
- *the total number of kicks to the kicker's natural side ('left' by convention) is smaller than the total number of jumps to the (kicker's) left*
- *if the natural side 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*
- *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*
- *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. Take for instance the prediction that the scoring probability should be the same on each side. The following simple counterexample shows that, even with two types of matches, the prediction does not hold on aggregate data. That is, although the scoring probabilities for R and L are the same for each match (but differ across matches), the total percentage of L kicks that are scored differs from that of R kicks.

Example 11 *Assume* $\mu = 0$ *(so that all equilibria are RR) and* $\pi_R = \pi_L = 1$ *(so that the goalie and the kicker play the same mixed strategy). There are two types of matches. In matches of type 1 (proportion k),* $P^R = .1$ *and* $P^L = .9$*; in matches of type 2 (proportion* $1 - k$*),* $P^R = P^L = .5$*. Then:*

- *for type 1 matches, the equilibrium strategy is* $l = \lambda = .9, r = \rho = .1$*, and the scoring probability is* $.91$ *on both sides*
- *for type 1 matches, the equilibrium strategy is* $l = \lambda = r = \rho = .5$*, and the scoring probability is* $.75$ *on both sides*

It follows that:

- *The total number of kicks to the left is* $.4k + .5$*, out of which* $.444k + .375$ *are scored, hence a proportion*

$$s_L = \frac{.444k + .375}{.4k + .5}$$

- *The total number of kicks to the right is* $-.4k + .5$*, out of which* $-.284k + .$

375 *are scored, hence a proportion*

$$s_R = \frac{-.284k + .375}{-.4k + .5}$$

It is easy to check that these proportions are never equal, except for $k = 0$ *or* 1 *(i.e., for homogeneous populations). In all other cases, one gets* $s_L > s_R$*.*

The intuition of this example is clear. A proportionally larger fraction of the kicks to the left comes from type 1 match, where the scoring probability is higher. Econometrically, this is exactly equivalent to a *selection bias*: the side of the kick is not exogenous but depends on the scoring probabilities.

In fact, any prediction entailing non linearities is generally not preserved by aggregation. Take the independence property discussed above, and assume, for expositional convenience, that μ is uniformly low, so that no GR equilibrium is observed. The overall probability of observing a kick to the right while the goalie jumps to the right is

$$\begin{aligned}
\Pr(R, L) &= \int r_0(P_R, P_L, \pi_R, \pi_L, \mu)\, \rho_0(P_R, P_L, \pi_R, \pi_L, \mu)\; d\phi(P_R, P_L, \pi_R, \pi_L, \mu) \\
&= \int \frac{(\pi_L - P_L)(\pi_R - P_L)}{(\pi_R + \pi_L - P_L - P_R)^2}\; d\phi(P_R, P_L, \pi_R, \pi_L, \mu)
\end{aligned}$$

whereas the product of the marginal probabilities is

$$\begin{aligned}
\int r_0 \rho_0 \, d\phi &= \left[\int \frac{\pi_L - P_L}{\pi_R + \pi_L - P_L - P_R}\; d\phi(P_R, P_L, \pi_R, \pi_L, \mu)\right] \\
&\quad \cdot \left[\int \frac{\pi_R - P_L}{\pi_L + \pi_R - P_L - P_R}\; d\phi(P_R, P_L, \pi_R, \pi_L, \mu)\right]
\end{aligned}$$

These two expressions need no longer be equal. Their difference is the correlation between the two events, as induced by the parameter distribution. That is, although for each match the two events are independent, the distribution of parameters may be such that those matches where the probability of kicking to the right is high are also, on average, those where the probability of jumping to the right is high. This is the case in particular whenever π_R and π_L are equal, since the kicker and the goalie then randomize between L and R with equal (conditional) probability. In that case, the difference can be interpreted as a variance, hence is always positive. The general message, here, is that although the independence property is satisfied for each match, heterogeneity tends to create correlation across individual behavior.

These remarks have important consequences in terms of empirical tests. First, our results indicate that, even for a given kicker and a given keeper, the scoring probabilities are affected by various exogenous variables. For instance, our empirical results suggest 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. Technically, this means that the game matrix varies with the covariates, which may introduce aggregation biases. Therefore, the equal scoring probability property should *not* be tested on raw data, but instead *conditional on observables* - which requires adequate econometric tools.[12] We shall consider this problem with some care in the empirical section.

[12] We find, however, that while scoring probabilities do change over time during the game, the probabilities of 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 second, and more disturbing problem is that while the total number of kicks available is fairly large, they mostly represent different matches of kickers and goalies. For any given match, 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-keeper pair faces the same matrix whatever the particular goalie at stake.

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 12 *Under assumption (IG), for any particular kicker i:*

- *The kicker's strategy does not depend on the goalkeeper*

- *The goalkeeper's strategy is identical for all goalkeepers*

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

- *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.*

- *Conditional on non 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. This drastically improves the power of our tests.

4 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.

Table II presents a breakdown of the number of penalty kicks per kicker in the data. A total of 162 kickers appear in the data. 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 4 shots (58% percent of the total observations) or the 9 kickers with at least 8 shots (22% percent of the total observations). Table II also presents the breakdown of penalty kicks per goalie. The number of goalies represented in the data (88) is much smaller than the number of shooters.

4.1 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[13]:

$$R_i^K = X_i\alpha + \beta R_i^G + \gamma \bar{R}_i^K + \delta \bar{R}_i^K + \varepsilon_i \qquad \text{(SM)}$$

where R_i^K (resp. R_i^G) is a dummy for whether, in observation i, the kicker shoots (resp. keeper jumps) right, R_i^{K*} is the corresponding latent variable, $\bar{R}_i^K$ (resp. $\bar{R}_i^K$) is the proportion of kicks by the kicker (resp. of jumps by the goalie) going right on all shots except this one, and 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 III. 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.[14] Columns 1 and 2 include only controls for the observed kicker and goalie behaviors. Columns

[13] Probit regressions give similar results, although the interpretation of the coefficients is less straightforward.

[14] In contrast to kickers, who may really have taken very few penalty kicks in their careers,

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 III 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, the goalie jumps 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 the wrong direction.[15]

A second observation that emerges from Table III 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.[16] 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.

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.

[15] As argued above, 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.

[16] 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.

4.2 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.[17] 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, Table IV presents information on the frequency with which we observe kickers and goalies who take the same action in each of the observations in our data. Focusing first on kickers in the first two columns of the table, there are no kickers in our sample with at least four kicks who always kick in one direction. Only 3 of the 26 kickers with exactly 3 penalty kicks always shoot the same direction. Even among kickers with exactly two shots, the same strategy is played both times in less than half the instances. Overall, there are 91 kickers in our sample with

[17] 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.

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 are observed always kicking the same direction, conditional on the number of kicks we observe for each kicker. This value, along with the appropriate standard error, is shown in the bottom row of the table.

Parallel evidence on whether goalies play mixed strategies is presented in columns 3 and 4 of Table IV. As with kickers, 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). Of the 69 goalies in the sample who face more than one shot, 13 are observed always jumping the same direction. Using the sample average frequencies of jumping left, center, or right, we would have predicted

Table V 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 three predictions of the model that can be tested using the information in Table V. First, the model predicts that the kicker will choose to play "center" more frequently than the goalie (this is the content of Proposition 3 above). This 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 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). Given independence, these results, in turn, imply that the cell "left-left" should have the greatest number of observations. This predictions is confirmed by the data, with the kicker and goalie both choosing left more than twenty-five percent of the time. The next most common outcome (goalie left, kicker right) appears about twenty percent of the time. Finally, the "right-right" pattern is the least frequent, as predicted by the model.

A third prediction of the model is that goalies should play "left" (the kicker's natural side) more frequently than kickers do. Indeed, goalies play "left" 260 times (56.6% of kicks), compared to 206 (44.9%) instances for kickers, a difference that is statistically significant at the .001 level. This result is somewhat exaggerated 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), but the difference is no longer significant.

For completeness, Table VI presents the matrix of scoring probabilities as a function of the actions taken by kickers and goalies. As noted 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. Kicking right has the lowest payoff, averaging only 70 percent success.

4.3 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 VII 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- 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 VI. The cutoff values for rejecting the null hypothesis at the .10 and .05 level respectively are 1.31 and 1.42. In none of the four cases can we reject the hypothesis that all goalies are identical.[18]

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

$$S_{i,t} = X_{i,t}\alpha + \sum \beta_i D_i + \sum \gamma_i D_i R_{i,t} + \sum \delta_i D_i M_{i,t} + \varepsilon_{i,t} \qquad \text{(SM)}$$

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, $R_{i,t}$ (resp. $M_{i,t}$) is a dummy for whether the kick goes right, and 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 $(\gamma_1, ..., \gamma_n, \delta_1, ..., \delta_n)$ are jointly insignificantly different from zero. The results of this test are presented in the top panel of Table VIII. 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 8 or more kicks in the sample (9 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 the game or the score of the game, then in principle these covariates should be

[18] 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 level in XXX of the four columns in Table VI.

included.[19] 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 .05 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 level in 5 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 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 more 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 VIII 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 VIII 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

[19] 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.

eight kicks. The structure of the bottom panel of 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 level.

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 (forthcoming) analysis of serves in tennis[20].

5 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

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

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. Arguably, the penalty kick data we examine more closely corroborates the predictions of theory than past laboratory experiments (cites) 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. We believe that the heterogeneity issue will be a necessary ingredient of future studies attempting to test game theory applications in real-world data.

Bibliography

Bresnahan, Timothy, and Peter Reiss, 1990, "Entry in Monopoly Markets," *Review of Economic Studies* 57:531-553.

Brown, J. and R. Rosenthal, 1990, "Testing the Minimax Hypothesis: A Re-examination of O'Neill's Game Experiment," *Econometrica* 58:1065-1081.

Hendricks, Kenneth, and Robert Porter, 1988, "An Empirical Study of an Auction with Asymmetric Information," *American Economic Review* 78:865-883.

McCabe, Kevin, Arijit Mukherji, and David Runkle, 2000, "An Experimental Study of Information and Mixed-Strategy Play in the Three-person matching-pennies Game," *Economic Theory* 15:421-462.

Mookherjee, Dilip, and Barry Sopher, 1994, "Learning Behavior in an Experimental Matching Pennies Game," *Games and Economic Behavior* 7:62-91.

Ochs, Jack, 1995, "Games with Unique, Mixed Strategy Equilibria: An Experimental Study," *Games and Economic Behavior* 10:202-217.

O'Neill, B., 1987, "Nonmetric Test of the Minimax Theory of Two-Person Zerosum Games," *Proceedings of the National Academy of Sciences* 84:2106-2109.

Rapoport, Amnon, and Richard Boebel, 1992, "Mixed Strategies in Strictly Competitive Games: A Further Test of the Minimax Hypothesis," *Games and Economic Behavior* 4:261-283.

Robson, A., 1994, "An 'Informationally Robust' Equilibrium in Two-Person Nonzero-Sum Games", *Games and Economic Behavior* 2:233-245.

Slonim, Robert, and Alvin Roth, 1998, "Learning in High Stakes Ultimatum Games: An Experiment in the Slovak Republic," *Econometrica* 66:569-596.

Walker, Mark, and John Wooders, forthcoming, "Minimax Play at Wimbledon," *American Economic Review.*