Nash Equilibrium and Penalty Kicks Outside the Top-5: It Is Not as Bad as You Think*

Egor Malkov†

June 24, 2018

Abstract

Is the behavior of people consistent with the predictions of the Nash equilibrium and, in particular, the Minimax hypothesis? The existing literature have not reached a consensus about the answer to this question. This paper studies whether the soccer players involved in a penalty kick behave themselves according to the predictions of the equilibrium play. Using a novel dataset consisting of 1679 penalty kicks, I show that the behavior of players is consistent with the Minimax hypothesis despite they represent a league that is of moderate quality in terms of skills. However, when I make use of the Kolmogorov-Smirnov test based on the randomized Fisher exact test — a test that is more powerful than any existing one used in the prior literature on the Minimax behavior — I call into question the result that success rates are identical across players' strategies, a necessary prediction of the Minimax hypothesis. This suggests that the existing positive results should be taken with a grain of salt.

Keywords: Nash equilibrium, minimax, mixed strategy, randomized test, soccer
JEL Classification Codes: C12, C72, D91, Z20

* I thank Beth Allen, David Rahman, Aldo Rustichini, and the participants of the Theory Workshop at the University of Minnesota for their comments. I thank FC Tom, in particular, Dmitry Shuba and Vasily Murtazin, for providing me with the penalty records for the 2007 season. The views expressed herein are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

† University of Minnesota and Federal Reserve Bank of Minneapolis. E-mail: `malko017@umn.edu`.

1 Introduction

Is the behavior of people consistent with the predictions of the Nash equilibrium and, in particular, the Minimax hypothesis? This question was challenged in the literature many times over the last several decades. Despite the seemingly easy theoretical concept, its empirical verification lacks the ultimate answer. Some papers argue that the answer is *yes*, while some disagree with that and conclude that it is *no*.

In this paper I join the aforementioned discussion and get a number of challenging results. The contribution of the paper is threefold. First, I collect a novel dataset comprised of 1679 penalty kicks that covers literally *all* the shots with the Russian Football Premier League (henceforth RFPL) players on one or another side (or both) of the penalty mark between the 2007 and 2017/18 seasons.

Second, I employ the tests that were used in the prior literature on Minimax behavior and show that the behavior of the players representing RFPL, a league of *lower quality* than the top-5 UEFA leagues, is consistent with equilibrium play. In fact, I show that (i) success rates are the same across strategies for all the players, and (ii) both kickers and goalkeepers' strategic choices are serially independent. These results hold when we consider different subgroups of players — kickers, goalkeepers, native players, and foreign players — as well. To the best of my knowledge, this is the first paper to show the consistency with the Minimax predictions for players representing a moderate-quality league. The previous attempt by Palacios-Huerta (2014) — for the U.S. Major League Soccer — shows that the players do not play Minimax there. This suggests that even in the moderate-quality leagues players can have Minimax strategic skills. It is not as bad as we think.

Finally, I make use of the test developed by Gauriot et al. (2018) that is more powerful than any existing one used in the prior literature that test the Minimax hypothesis. I believe, this paper is the first one that performs this test on the penalty kicks data. I show that on the aggregate level the result that success rates are the same across strategies for all the players has to be overturned. In addition, it has to be overturned for the subgroups of goalkeepers and native players as well. However, it still holds for the subgroups of strikers and foreign players. I show that the main driver of the failures is the behavior of native goalkeepers. Again, it is not as bad as we think.

The reasons for choosing RFPL are multiple. First, in the UEFA rankings for club competitions Russia holds the sixth place[1], right after the top-5 leagues. On the one hand, it is interesting to test the behavior of kickers and goalkeepers who play in the league ranked right after the top-5. However, this seeming relative strength of the league is quite questionable. Since 2007 only four Russian players — Andrey Arshavin ("Arsenal", 2009-2013), Diniyar Bilyaletdinov ("Everton", 2009-2012), Roman Pavlyuchenko ("Tottenham Hotspur", 2008-2012), and Yuri Zhirkov ("Chelsea", 2009-2011) — joined the clubs representing the top-5 UEFA leagues. The Russian national team, consisting almost fully from the players representing RFPL, cannot be treated as a strong team in the period of 2007-2018.[2] One possible explanation for this discrepancy between the high rank of Russia in the UEFA rankings for club competitions and its weak performance at the level of national teams is that the foreign players in RFPL potentially are more qualified than the native players. I study this argument by separating the foreign and native players into different subgroups and testing whether the behavior of each group is consistent with Minimax. I have strong reasons to consider at least the native players in RFPL as those with moderate quality of skills. It is interesting to test the behavior of moderate-quality players and compare the results with Palacios-Huerta (2014). Finally, RFPL is a highly competitive league among the top-7 UEFA soccer leagues, (see Appendix A for a short discussion). This means that the potential importance of each penalty kick and the associated stakes there are higher than in less competitive leagues.

As I argued at the very beginning, there is no consensus in the literature about the the question of whether people's behavior is consistent with the predictions of equilibrium play. The studies based on the experimental data generally reject the hypothesis about the Minimax behavior. One of the pioneering papers, O'Neill (1987), argues that the results of experiments considered there are consistent with the Nash equilibrium. However, later these conclusions were contested by a number of authors, including Brown and Rosenthal (1990), Walker and Wooders (2001) and some others. Palacios-Huerta and Volij (2008) show that professional soccer players do play Minimax not only in the field, but also in a laboratory environment, i.e. they are able to transfer the skills from the field to the laboratory. For

[1] https://www.uefa.com/memberassociations/uefarankings/country/index.html. Accessed on June 1, 2018.

[2] The only striking exclusion was Euro 2008 where the Russian national team won the bronze. However, this period corresponds to the beginning of my sample.

comparison, they also show that the behavior of college students, who are not considered as professionals, is far from the Nash equilibrium predictions. However, these results are reexamined in Wooders (2010) who uses the data from Palacios-Huerta and Volij (2008) and conclude that neither the professionals, nor the students do demonstrate the behavior that is consistent with the Minimax hypothesis. Another paper, Levitt et al. (2010), finds little evidence that real-world experience is transferred to the laboratory setting.

Another strand of the literature, which I follow in this paper, study the behavior of people in a natural setting. The pioneering work in this area is Walker and Wooders (2001) who study the behavior of tennis players serving the ball. They show that the first prediction of the Minimax hypothesis — player's winning probabilities has to be the same for left and right serves — holds in the data, while the second — serial independence of strategic choices — fails to hold. Even the best tennis players tend to switch from one strategy to another too often. A complementary paper by Kovash and Levitt (2009) shows that baseball and (American) football players also fail to randomize over their actions. Hsu et al. (2007) extend the sample from Walker and Wooders (2001) and show that both Minimax predictions hold in the data. However, recently Gauriot et al. (2018) have overturned some of their results.

Perhaps, the most popular way to study the implications of the Minimax hypothesis in a natural setting in the literature is to look at the interaction between a kicker and a goalkeeper in a penalty kick. Two pioneering papers in this area are Chiappori et al. (2002) and Palacios-Huerta (2003). Both show that the behavior of soccer players is consistent with the predictions of equilibrium play. I will discuss all the related details throughout the paper. Other papers who study the behavior of soccer players in the light of game theory predictions include Moschini (2004), Coloma (2007), Baumann et al. (2011), Misirlisoy and Haggard (2014), and Dohmen and Sonnabend (2018). Azar and Bar-Eli (2011) show that the mixed-strategy Nash equilibrium under the assumption of simultaneity of moves describes the behavior of players in penalty kicks better than many other alternative models. Bar-Eli et al. (2007) argue that goalkeepers are subject to the omission bias, a bias in favor of action, and therefore jump right or left more frequently than stay at the center.

The paper is organized as follows. Section 2 describes a penalty kick through the lens of a 2-by-2 game. Section 3 describes the data. Section 4 explains the empirical results. In Section 5 I make use of the test developed by Gauriot et al. (2018). Section 6 concludes.

2 Nash Equilibrium and Penalty Kicks

Professional soccer and, in particular, penalty kicks serve as a suitable playground for the researchers who study game theory predictions in the data. The obvious advantages are the following: the rules that regulate penalty kicks are quite simple and transparent[3], players have specific professional skills, the stakes are pretty high, the actions of both a kicker and a goalkeeper are observable, players move simultaneously[4], the outcomes are immediately observable, and the payoffs are easy to describe.

Any normal form game is described by three elements:

1. *The set of players* $\mathcal{I}$. In the case of a penalty kick,

$$\mathcal{I} = \{\text{kicker, goalkeeper}\}.$$

2. *The pure-strategy space* $\mathcal{S}^i$ *for each player* $i \in \mathcal{I}$. In the case of a penalty kick[5],

$$\mathcal{S}^K = \mathcal{S}^G = \{\text{natural side (N), non-natural side (NN)}\}.$$

3. *The payoff functions* u^i. Following the literature, e.g., Palacios-Huerta (2003), I assume that in the case of a penalty kick, $u^K(s)$ and $u^G(s)$, where $s = (s_K, s_G)$ is the profile of strategies, are defined as the probabilities of success. Thus, since a kicker's success is a goalkeeper's failure and vice versa, this is a constant-sum game. In this game the Minimax solution coincides with the Nash equilibrium.

Having described the game, I draw the matrix of it. Since we deal with a constant-sum game, it is completely determined by the kicker's probabilities of success. Note that we do not

[3] See FIFA (2017) for the most recent soccer rules.

[4] As Palacios-Huerta (2003) points out, usually it takes 0.3 seconds for the ball to travel to the goal line after the kick. This time is not enough for the goalkeeper to react and move accordingly to stop the ball. Later I will show that the hypothesis about simultaneity of moves is not rejected in the data.

[5] The natural side for the right-footed kicker is the left-hand side of the kicker (the right-hand side of the goalkeeper) and the right-hand side for the left-footed kicker (the left-hand side of the goalkeeper). The non-natural side is defined in the opposite way. I will discuss this terminology more thoroughly later in Section 3.

observe the true payoffs as, for instance, the players in the laboratory from Palacios-Huerta and Volij (2008) who use the clearly specified matrices of the game.

		Goalkeeper	
		N	NN
Kicker	N	$\pi_{N,N}$	$\pi_{N,NN}$
	NN	$\pi_{NN,N}$	$\pi_{NN,NN}$

A unique mixed strategy Nash equilibrium of this game exists when

$$\pi_{N,NN} > \pi_{N,N} < \pi_{NN,N}$$

$$\pi_{NN,N} > \pi_{NN,NN} < \pi_{N,NN}$$

If a penalty kick is represented by the game described above, then both the kickers and goalkeepers must use mixed strategies in the Nash equilibrium. Then, the theory provides us with two predictions that can be tested with the data:

1. Success rates should be the same across strategies for both the kickers and goalkeepers. In particular,

$$p_N^K = g_N \pi_{N,N} + (1 - g_N)\pi_{N,NN}$$

$$p_{NN}^K = g_N \pi_{NN,N} + (1 - g_N)\pi_{NN,NN}$$

$$p_N^G = k_N(1 - \pi_{N,N}) + (1 - k_N)(1 - \pi_{NN,N})$$

$$p_{NN}^G = k_N(1 - \pi_{N,NN}) + (1 - k_N)(1 - \pi_{NN,NN})$$

   where p_N^K and p_{NN}^K are the probabilities of the kicker's success when choosing the natural and non-natural sides respectively, p_N^G and p_{NN}^G are the probabilities of the goalkeeper's success, k_N is the kicker's probability of the natural side choice, g_N is the goalkeeper's probability of the natural side choice. In equilibrium, $p_N^K = p_{NN}^K$ and $p_N^G = p_{NN}^G$.

2. Both the kicker and goalkeeper's strategic choices must be serially independent.

Before turning to the results of formal tests for these two predictions, I firstly describe the data and provide some descriptive statistics.

3 Data

I use a unique dataset comprised of 1679 penalty kicks with at least one opponent representing RFPL between the 2007 and 2017/18 seasons.[6] These observations come from RFPL itself, Russian Cup, UEFA competitions, such as the UEFA Champions League, the UEFA Europa League[7], and UEFA Intertoto Cup, games for national teams, and RFPL relegation games. Thus, the dataset covers literally *all* the penalty kicks with RFPL players on one or another side (or both) of the penalty mark.

To collect the data I have watched the TV program *Futbol Rossii*, soccer news, official highlights of the games prepared by the soccer clubs, or, when the highlights, news or programs were not available or a game was beyond the scope of them, the records of the games themselves.

The data includes the names of the kicker and goalkeeper for each penalty kick, the names of the teams they represent, the choices of the players (left, center, or right), the nationalities of the players (native or foreign), the date of the game, the final score of the game, the score right before the penalty kick, the minute when the penalty kick is performed, the foot the kicker use to kick the ball, and the outcome of the penalty kick (scored or not).[8] If the goal from the penalty kick is not scored, the data indicates whether the ball was saved by the goalkeeper, or hits the goalpost, or there is a wide shot. Finally, for penalty shootouts the dataset includes the corresponding ordering numbers of the penalty kicks.

Table 1 shows the summary statistics for the left-footed and right-footed kickers separately.[9] I intentionally omit the table where the left-footed and right-footed kickers are

[6] RFPL is contested by 16 teams. The two lowest-placed teams are relegated to the Russian Football National League (RFNL), the second-level league of Russian professional soccer. The teams that finish 13^{th} and 14^{th} play two relegation games against the teams that finish 4^{th} and 3^{rd} respectively in RFNL. Before the 2011 a typical RFPL season lasted from spring to fall. However, since the 2011/12 season it lasts from summer to spring.

[7] Until the 2009/10 season the UEFA Europa League was known as the UEFA Cup.

[8] I use the data on the first shots only. If the goal is not scored from the first shot, I treat this as *not scored*.

[9] There is only one player in the sample — Alexey Sapogov — who shot one penalty kick using his left foot

treated as homogeneous since it is an established fact that the right-footed kickers tend to kick to the left more frequently than the left-footed kickers, while the left-footed kickers — to the right (see Chiappori et al. (2002) or Palacios-Huerta (2003)). The right-footed kickers represent 79.1 percent of the sample, while the fraction of penalty kicks performed by the right-footed kickers is 79.4 percent.

First, Table 1 shows the frequencies of different strategy profiles. Each strategy profile ij consists of the kicker's strategy i (kick to the right, R, left, L, or center, C) and the goalkeeper's strategy j (jump to the right, R, left, L, or stay at the center, C). The first letter denotes the kicker's choice, the second one — the goalkeeper's choice. The choices are taken from the kicker's perspective. For instance, "LL" means that the kicker shoots to the left, while the goalkeeper jumps to his right. Second, I condition the scoring rates on different strategy profiles and score differences right before the penalty kick. Score differences are taken from the kicker's perspective. For instance, "-1" means that the kicker's team is behind one goal right before the penalty kick.

There are several immediate observations. First, the scoring rate for the left-footed kickers is slightly higher than for the right-footed kickers — 77.5 percent against 73.4 percent.[10] This is consistent with the evidence from Baumann et al. (2011) who study the German Bundesliga in the period of 1995-2007. Second, the right-footed kickers do shoot to the left more frequently than the left-footed kickers (48.5 percent against 40.2 percent). Similarly, the left-footed players tend to kick more to the right (48.6 percent against 34.8 percent). Third, the scoring rates of the kickers whose teams are behind one or more goals are typically lower than for those who are leading. One possible explanation is that the team that is losing at the time of the penalty kick may be weaker than the opponent, and, following the idea of sorting, the kicker from this team may have weaker skills and, therefore, lower scoring rate. This observation is somewhat opposite to Palomino et al. (2000) who show that, everything else being equal, a team that is down by a goal is, in equilibrium, more likely to score than a team in any other competitive position.

and another one — using his right foot. All the other kickers use one foot only.

[10] In fact, the difference is statistically significant at the 10% significance level. Under the null hypothesis that scoring rates are equal and the alternative one that the scoring rate of the left-footed kickers is higher than for the right-footed kickers, t-test gives the p-value of 0.054.

Table 1: Frequency of strategy profiles and scoring rates by right- and left-footed kickers

Score difference	Obs.	LL	LC	LR	CL	CC	CR	RL	RC	RR	Scoring rate, %
Right-footed kickers											
$\leq$ -2	86	20.9	0.0	20.9	5.8	3.5	11.6	23.3	0.0	14.0	68.6
-1	178	34.8	0.0	18.5	9.0	0.6	6.7	12.9	0.6	16.9	70.8
0	406	29.8	0.5	17.0	10.8	1.2	6.2	18.2	0.5	15.8	76.1
1	166	32.5	0.6	21.1	6.0	1.8	4.8	21.1	0.6	11.4	78.3
$\geq$ 2	67	29.9	1.5	13.4	9.0	0.0	11.9	16.4	3.0	14.9	74.6
Shootout	430	26.5	2.3	18.1	7.9	2.8	5.1	17.9	1.4	17.9	70.9
Total	1333	29.2	1.1	18.2	8.6	1.8	6.4	18.0	0.9	15.9	
Scoring rate, %	73.4	59.9	85.7	93.8	80.9	12.5	75.3	96.3	91.7	49.5	
Left-footed kickers											
$\leq$ -2	24	20.8	0.0	29.2	0.0	0.0	0.0	16.7	0.0	33.3	79.2
-1	52	13.5	0.0	26.9	3.8	5.8	1.9	17.3	0.0	30.8	71.2
0	108	21.3	0.0	23.1	5.6	0.0	8.3	20.4	0.9	20.4	78.7
1	43	14.0	0.0	18.6	4.7	0.0	7.0	18.6	0.0	37.2	76.7
$\geq$ 2	20	15.0	5.0	10.0	0.0	5.0	10.0	25.0	0.0	30.0	90.0
Shootout	99	18.2	1.0	19.2	1.0	3.0	6.1	23.2	2.0	26.3	76.8
Total	346	17.9	0.6	21.7	3.2	2.0	6.1	20.5	0.9	27.2	
Scoring rate, %	77.5	59.7	100.0	97.3	81.8	0.0	81.0	94.4	100.0	63.8	

Notes: Each strategy profile is denoted by ij with i being the kicker's strategy, j — the goalkeeper's strategy. The sides are taken from the kicker's perspective, i.e. "L"("R") denoted the left-hand side (right-hand side) of the kicker and the right-hand side (left-hand side) of a goalkeeper. "C" denotes the center. Percentages may not add up to 100% due to rounding.

Following the previous studies that use the data on penalty kicks as a ground for testing the Minimax hypothesis, I consider the "natural" and "non-natural" directions of the kickers. The natural side for the right-footed kickers is "L" (left from the kicker's perspective, right from the goalkeeper's perspective), while for the left-footed kickers — "R" (right from the kicker's perspective, left from the goalkeeper's perspective). On the other hand, the non-natural side for the right-footed kickers is right, while for the left-footed kickers — left. By now I do not treat the choice "C" as the natural one. Palacios-Huerta (2003) and Palacios-Huerta (2014) assume that "C" is included into the natural direction. Later, when I consider the individual behavior of players, I will use this assumption as well.

To justify the use of the natural and non-natural sides, I test the hypothesis that the game is identical for the right-footed and left-footed kickers up to the renaming of their choices. I run three regressions with the following dependent variables — (i) the kicker shoots to the natural side, (ii) the goalkeeper jumps to the natural side, and (iii) the goal is scored. The right-hand side of these regressions includes the kicker-type fixed effects (right-footed and left-footed) and a set of covariates, including six dummy variables corresponding to 15-minute intervals of the game, five dummy variables corresponding to the score differences right before the penalty kick as in Table 1, whether the kicker's team plays at home, whether the game is played on the neutral field, whether the penalty kick is shot in the extra time, whether the penalty kick is performed in the shootout, the interaction terms that absorb any systematic differences in outcomes across years within a competition or across competitions, and, finally, the goalkeeper-fixed effects. In all three cases I do not reject the null hypothesis that the kicker-type fixed effects are jointly insignificantly different from zero. The p-values of the F-statistic are (i) 0.730, (ii) 0.802, and (iii) 0.419.

Table 2 provides the descriptive statistics for the frequencies of strategy profiles and scoring rates using the natural and non-natural sides terminology. As expected, kickers tend to choose the natural side more frequently than center or the non-natural side (48.4, 15.6, and 35.9 percent respectively). The total scoring rate is 74.3 percent which means that, in average, every fourth penalty kick is not scored. This number is close to one from Chiappori et al. (2002), who get 74.9 percent, and somewhat lower than 80.1 percent from Palacios-Huerta (2003) or Palacios-Huerta (2014). As in Table 1, the argument that the team that is behind at the time of the penalty kick has lower chances to score is supported again. Also note that the scoring rates are quite close across the first and second halves of the game. Finally, in the shootout the kickers have slightly lower than average chances of scoring.

The average number of goals per game in the sample is 3.24.[11] The fraction of cases with -1, 0, or 1 difference at the time of the penalty kick is almost 83 percent. Thus, the outcome of any given penalty kick is often plays a crucial role for the final result of the game.

[11] This is greater than 2.4, the average number of goals scored in all the RFPL games between the 2007 and 2017/18 seasons. When I constrain my sample to RFPL games only, the average number of goals is 3.16. The median is 3.

Table 2: Frequency of strategy profiles and scoring rates

	Obs.	N-N	N-C	N-NN	C-N	C-C	C-NN	NN-N	NN-C	NN-NN	Scoring rate, %
Score difference											
$\leq$ -2	110	23.6	0.0	20.0	4.5	2.7	9.1	24.5	0.0	15.5	70.9
-1	230	33.9	0.0	18.3	7.4	1.7	6.1	16.1	0.4	16.1	70.9
0	514	27.8	0.6	17.7	10.3	1.0	6.0	19.3	0.4	16.9	76.7
1	209	33.5	0.5	20.6	6.2	1.4	4.8	20.6	0.5	12.0	78.0
$\geq$ 2	87	29.9	1.1	16.1	9.2	1.1	9.2	14.9	3.4	14.9	78.2
1st half	480	31.3	0.4	17.9	7.9	1.0	5.6	19.8	0.4	15.6	75.6
2nd half	662	28.9	0.5	18.9	8.6	1.7	6.8	18.4	0.8	15.6	75.2
Extra-time	94	37.2	1.1	10.6	8.5	1.1	4.3	18.1	0.0	19.1	74.5
Last 10 min	207	31.4	0.5	17.9	7.7	2.4	6.3	15.5	1.4	16.9	74.4
Shootout	529	26.5	2.3	19.1	7.6	2.8	4.3	18.1	1.3	18.0	72.0
Total	1679	28.8	1.0	18.6	8.1	1.8	5.7	18.8	0.8	16.3	
Scoring rate, %	74.3	60.7	88.2	93.9	80.9	9.7	76.0	96.5	92.9	51.8	

Notes: Each strategy profile is denoted by ij with i being the kicker's strategy, j — the goalkeeper's strategy. The sides are taken from the kicker's perspective. "N"("NN") denotes the natural side (non-natural side). "C" denotes the center. Percentages may not add up to 100% due to rounding.

4 Empirical Results

The main purpose of this section is to test two predictions of the Minimax hypothesis:

1. Success rates should be the same across strategies for both the kickers and goalkeepers.

2. Both the kicker and goalkeeper's strategic choices must be serially independent.

The success rate for the kicker is defined as the number of goals scored from penalty kicks divided by the number of penalty kicks he performed. On the contrary, the success rate for the goalkeeper is defined as the number of penalty kicks against him that were not scored[12] divided by the number of penalty kicks he was involved in.

[12] The goal is not scored when either the goalkeeper saves the ball, or the ball hits the goalpost, or there is a wide shot.

In this section I employ the tests that were previously used in the literature studying the consistency of the Nash equilibrium predictions with either the experimental or field data. In the next section I make use of the test developed by Gauriot et al. (2018), the most powerful test of the Minimax hypothesis ever used in the literature.

Before addressing the issues of consistency, first of all, I check the validity of the fundamental assumption that the kickers and goalkeepers move simultaneously. I follow the approach proposed by Chiappori et al. (2002) and run the following linear probability regression:

$$N_i^K = \mathbf{X}_i \alpha + \beta N_i^G + \gamma \bar{N}_i^K + \delta \bar{N}_i^G + \varepsilon_i \tag{1}$$

where N_i^K is the dummy for whether the kicker chooses the natural side, N_i^G is the dummy for whether the goalkeeper jumps to the natural side, $\bar{N}_i^K$ is the fraction of shots to the natural side *before* the given penalty kick, $\bar{N}_i^G$ is the fraction of jumps to the natural side *before* the given penalty kick, $\mathbf{X}_i$ is the set of controls (for their description see Section 3 where I show that the right-footed and left-footed kickers are identical up to renaming of their choices). I consider this specification — with $\bar{N}_i^K$ and $\bar{N}_i^G$ being calculated on the basis of the past choices only — to be easier for interpretation rather than alternative one where $\bar{N}_i^K$ and $\bar{N}_i^G$ are calculated on the basis of all kicks except the one at stake.[13]

Estimation results are presented in Table 3. In columns (2) and (4) I restrict the sample to the kickers with at least 5 kicks. In all the four cases I do not reject the null hypothesis that $\beta = 0$. Thus, the action of the goalkeeper in a given penalty kick does not predict the action of the kicker in this penalty kick.

Furthermore, we can see that the kickers who previously shot to the natural side more frequently are more likely to choose the natural side in a given penalty kick. For instance, for the full sample of kickers with controls included, additional 10 percent in the fraction of prior shots to the natural side leads to 2.5 percent more frequent shots to the natural side in the current penalty kick. The size of coefficient $\hat{\gamma}$ under different specifications from Table 3 and Table B.1 is of similar magnitude with one from Chiappori et al. (2002).

[13] See Table B.1 in Appendix B for the results under this alternative specification. In this case I do not reject the hypothesis about simultaneity of the kicker and goalkeeper's actions as well. Pierre-André Chiappori, Steven Levitt, and Timothy Groseclose use this specification in Chiappori et al. (2002).

Table 3: Simultaneity of the kicker and goalkeeper's actions

	Dependent variable: Kicker chooses the natural side			
	(1)	(2)	(3)	(4)
Goalkeeper chooses the natural side	0.0279 (0.0356)	0.0147 (0.0427)	0.0237 (0.0360)	0.0083 (0.0435)
Fraction of natural-side kicks *before* given penalty kick	0.2599*** (0.0503)	0.1951*** (0.0705)	0.2543*** (0.0505)	0.2008*** (0.0707)
Fraction of natural-side jumps *before* given penalty kick	0.0537 (0.0750)	-0.0127 (0.0879)	0.0611 (0.0756)	-0.0033 (0.0888)
Year × competition dummies	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes
Kickers with at least 5 kicks	No	Yes	No	Yes
Adj. R^2	0.0292	0.0063	0.0315	0.0111
Number of observations	852	618	852	618

Notes: *** $p < 0.01$. *Standard errors are in parentheses.*

Finally, Table 3 shows that the kickers do not take into account the history of the goalkeepers' actions when they shoot the penalty kicks. In all the columns I do not reject the hypothesis that $\delta = 0$. In other words, the kickers treat the goalkeepers as if they are all identical. This observation is consistent with the results from Chiappori et al. (2002). Note, however, that it is not possible to distinguish between two possible interpretations: either the kickers have the information about the prior actions of the goalkeepers but do not use it or, alternatively, the kickers do not have this information.[14] In addition, I also run four regressions with the following dependent variables: (i) the kicker shoots to the natural side, (ii) the kicker shoots to the center, (iii) the goalkeeper jumps to the natural side, and (iv) the goal is scored. The set of covariates consists of all the standard controls, discussed above, and the kicker-fixed effects. The p-values of the F-statistic for the joint significance of the goalkeeper-fixed effects are (i) 0.944, (ii) 0.394, (iii) 0.071, and (iv) 0.934. The p-values of the F-statistic for the joint significance of the kicker-fixed effects are (i) 0.024, (ii) 0.025,

[14] When I run regressions with dummy N_i^G being a dependent variable, I do not reject $\beta = 0$ and $\delta = 0$ and show that the goalkeepers are more likely to jump to the natural side in a given penalty kick when face kickers who previously shot to the natural side more frequently.

(iii) 0.634, and (iv) 0.034. Thus, at the 5% significance level there is substantial evidence of homogeneity across the goalkeepers and heterogeneity across the kickers.

The last thing I do before turning to the tests is the comparison of the observed frequencies with ones implied by the Nash equilibrium.[15] The 2-by-2 representation of a penalty kick with observed scoring probabilities is

		Goalkeeper	
		g_N	$1-g_N$
Kicker	k_N	60.66	93.93
	$1-k_N$	96.51	51.82

with k_N denoting the frequency of choosing the natural side by the kicker, g_N — the frequency of choosing the natural side by the goalkeeper. Table 4 shows that the observed and predicted frequencies are quite close to each other.

Table 4: Observed and predicted frequencies

	k_N, %	$1-k_N$, %	g_N, %	$1-g_N$, %
Observed frequencies	57.47	42.53	57.62	42.38
Predicted frequencies	57.32	42.68	54.02	45.98

We turn to the formal tests with this optimistic evidence at hand.

4.1 Individual Behavior

To allow for more observations now and thereafter I use the assumption proposed by Palacios-Huerta (2003) and treat "C" as the choice of the natural side.[16] I conduct the tests on the sample consisting of 15 kickers[17] with at least 15 penalty kicks and 26 goalkeepers with at least 20 penalty kicks. Given the discussion about possible differences between the native

[15] To calculate predicted frequencies, I use the equilibrium conditions stated in Section 2.

[16] As in Section 3, I test the null hypothesis that the right-footed and left-footed kickers are identical up to the renaming of their choices under a new definition of the natural side. The p-values of the F-statistic are (i) 0.521, (ii) 0.342, and (iii) 0.288. To check the robustness, I redo all the tests for the sample with discarded "C" choices. The results are quite similar and are available by request.

[17] In Section 4 and Section 5 I also add 9 penalty kicks performed by Roman Pavlyuchenko, the only Russian in our restricted sample who has experience in the top-5 league, when he played for "Tottenham Hotspur" in 2008-2012. When I add these observations, the conclusions do not change at all.

and foreign players, I also keep track of nationalities of the players so that I can test the hypotheses about their behavior. I define a player to be native if he was born in Russia. I define a player to be foreign if he was born outside of Russia. While this classification works in most of cases, there are several instances with ambiguity.[18] However, the results of the paper are robust even if we reshuffle the nationalities of the players that are under question.

4.1.1 Equality of Success Rates

To test the hypothesis that the success rates are similar across strategies for both the kickers and goalkeepers, I use Pearson's chi-square goodness-of-fit test for equality of two distributions.

Denote the success rate (alternatively, success probability) of player i who chooses strategy $j \in \{N, NN\}$ by p^i_j. Under the null hypothesis, $p^i_N = p^i_{NN} = p^i$. The Pearson statistic for player i is

$$Q^i = \sum_{j \in N,NN} \left[\frac{(S^i_j - n^i_j p^i)^2}{n^i_j p^i} + \frac{(F^i_j - n^i_j(1-p^i))^2}{n^i_j(1-p^i)} \right] \tag{2}$$

where n^i_j is the number of times player i chooses strategy j, S^i_j (F^i_j) is the number of successful (unsuccessful) times when player i chooses strategy j. Note that a scored goal is a successful (unsuccessful) outcome for the kicker (goalkeeper), and when a goal is not scored it is an unsuccessful (successful) outcome for the kicker (goalkeeper). Since I use the maximum likelihood estimate $\tilde{p}^i = (S^i_N + S^i_{NN})/(n^i_N + n^i_{NN})$, rather than using known p^i, Q^i is distributed asymptotically as chi-square with one degree of freedom.

Table 5 contains the results of Pearson tests. First, there is one rejection at the 5% significance level (Georgi Peev, kicker) and three rejections at the 10% significance level (Yaroslav Hodzyur and Artem Rebrov, both goalkeepers, in addition to Peev). With 41 observations at hand, the expected number of rejections at 5% is 2.05, at 10% — 4.1. The number of actual rejections is below the number predicted by the theory. In a nutshell, the hypothesis that the success rates are similar across strategies for each individual player is not rejected for most of the players representing RFPL at conventional significance levels.

[18] For instance, Andrey Karyaka was born in Dnipropetrovsk, Ukraine, however, never played for the Ukrainian national team, but for the Russian national team.

Table 5: Tests for equality of success rates

Player	Nat.	Foot	Obs.	Side choice N	Side choice NN	Success rate N	Success rate NN	Pearson	P-value
Kickers									
Roman Pavlyuchenko	NAT	Right	25	68.0	32.0	76.5	75.0	0.006	0.936
Hulk	FOR	Left	27	51.9	48.1	78.6	76.9	0.011	0.918
Bibras Natkho	FOR	Right	38	55.3	44.7	81.0	82.4	0.012	0.912
Aleksandr Samedov	NAT	Right	17	47.1	52.9	75.0	77.8	0.018	0.893
Roman Adamov	NAT	Right	16	43.8	56.2	71.4	77.8	0.085	0.771
Andrey Karyaka	FOR	Right	16	50.0	50.0	87.5	75.0	0.410	0.522
Sergey Kornilenko	FOR	Left	17	58.8	41.2	70.0	85.7	0.565	0.452
Seydou Doumbia	FOR	Right	15	53.3	46.7	87.5	71.4	0.603	0.438
Dmitry Sychev	NAT	Right	15	60.0	40.0	88.9	100.0	0.714	0.398
Aleksandr Kerzhakov	NAT	Right	21	71.4	28.6	86.7	100.0	0.884	0.347
Artem Dzyuba	NAT	Right	24	70.8	29.2	76.5	57.1	0.897	0.344
Dmitry Kombarov	NAT	Left	24	37.5	62.5	77.8	93.3	1.244	0.265
Fedor Smolov	NAT	Right	15	80.0	20.0	91.7	66.7	1.298	0.255
Vágner Love	FOR	Right	21	57.1	42.9	66.7	88.9	1.400	0.237
Georgi Peev	FOR	Right	21	66.7	33.3	100.0	71.4	4.421	0.035**
Goalkeepers									
Sergey Pareiko	FOR		29	55.2	44.8	31.3	30.8	0.001	0.978
Evgeny Gorodov	NAT		26	42.3	57.7	27.3	26.7	0.001	0.973
Ilya Abaev	NAT		23	73.9	26.1	17.6	16.7	0.003	0.957
Anton Shunin	NAT		23	65.2	34.8	26.7	25.0	0.008	0.931
Sergey Ryzhikov	NAT		52	67.3	32.7	31.4	29.4	0.022	0.882
Artur Nigmatullin	NAT		25	68.0	32.0	29.4	25.0	0.053	0.819
Soslan Dzhanaev	NAT		30	43.3	56.7	30.8	23.5	0.197	0.657
Mikhail Kerzhakov	NAT		28	64.3	35.7	38.9	30.0	0.221	0.638
Marinato Guilherme	FOR		49	57.1	42.9	21.4	28.6	0.331	0.565
Anton Kochenkov	FOR		28	71.4	28.6	15.0	25.0	0.390	0.533
Sergey Pesyakov	NAT		36	69.4	30.6	28.0	18.2	0.393	0.531
Aleksandr Selikhov	NAT		28	67.9	32.1	21.1	11.1	0.412	0.521
Stipe Pletikosa	FOR		58	63.8	36.2	32.4	23.8	0.481	0.488
Igor Akinfeev	NAT		75	48.0	52.0	36.1	28.2	0.538	0.463
Andrey Sinitsyn	NAT		20	55.0	45.0	18.2	33.3	0.606	0.436
Andriy Dykan	FOR		27	66.7	33.3	38.9	22.2	0.750	0.386
Yuri Lodygin	NAT		22	63.6	36.4	57.1	37.5	0.786	0.375
David Yurchenko	FOR		35	57.1	42.9	20.0	33.3	0.798	0.372
Aleksandr Belenov	NAT		51	43.1	56.9	45.5	31.0	1.113	0.291
Vyacheslav Malafeev	NAT		39	56.4	43.6	40.9	23.5	1.304	0.254
Dmitry Khomich	NAT		31	45.2	54.8	42.9	23.5	1.312	0.252
Vladimir Gabulov	NAT		52	71.2	28.8	21.6	6.7	1.668	0.197
Sergey Narubin	NAT		40	70.0	30.0	28.6	8.3	1.973	0.160
Yuri Zhevnov	FOR		27	70.4	29.6	26.3	0.0	2.584	0.108
Yaroslav Hodzyur	FOR		22	54.5	45.5	41.7	10.0	2.758	0.097*
Artem Rebrov	NAT		32	53.1	46.9	41.2	13.3	3.056	0.080*

Notes: "NAT" denotes a native player, "FOR" denotes a foreign player.

4.1.2 Serial Independence of Side Choices

To test the hypothesis of serial independence of side choices, I perform several tests. First, I use the runs test. Second, I make use of the approach from Arellano and Carrasco (2003). In addition, I estimate the serial correlation of players' choices.

Existing literature on randomization mostly rejects the hypothesis that people are able to generate random sequences: they switch their strategies either too often or too rarely. Walker and Wooders (2001) show that while the behavior of professional tennis players is consistent with the equal winning probabilities prediction, it fails to satisfy the serial independence: even the best tennis players tend to switch actions too often, i.e. demonstrate the negative serial correlation. Hsu et al. (2007) expanded the dataset from Walker and Wooders (2001) and show that the professional tennis players do randomize over the actions. In turn, Kovash and Levitt (2009) show that the baseball and (American) football players fail to randomize over their actions. Finally, Palacios-Huerta (2014) conclude that most players from the top soccer leagues are able to randomize over the choices of sides in penalty kicks, while the players from weaker leagues, such as MLS, fail the randomization tests.

To test the randomization hypothesis, I firstly address to the runs test. For this purpose, I order the penalty kicks by the time of their occurrence for each player and define a *run* as a succession of one or more particular side choices which are followed and preceded by a different side choice or no side choice at all.[19] Denote the total number of runs in player i's record by r^i. Then, denote by n^i_N and n^i_{NN} the numbers of natural and non-natural side choices respectively. The total number of player i's observations is $n^i = n^i_N + n^i_{NN}$.

By Theorem 2.2 from Gibbons and Chakraborti (2003), the probability distribution of r^i is given by

$$f(r^i|n^i_N, n^i_{NN}) = \begin{cases} 2\binom{n^i_N-1}{r/2-1}\binom{n^i_{NN}-1}{r/2-1}\Big/\binom{n^i_N+n^i_{NN}}{n^i_N}, & \text{if } r^i \text{ is even} \\ \left(\binom{n^i_N-1}{(r-1)/2}\binom{n^i_{NN}-1}{(r-3)/2}+\binom{n^i_N-1}{(r-3)/2}\binom{n^i_{NN}-1}{(r-1)/2}\right)\Big/\binom{n^i_N+n^i_{NN}}{n^i_N}, & \text{if } r^i \text{ is odd} \end{cases} \tag{3}$$

for $r^i = 2, 3, ..., n^i$.

[19] For instance, the sequence N, N, NN, N, NN, NN contains four runs.

Furthermore, Gibbons and Chakraborti (2003) show that when $n^i_N > 12$ and $n^i_{NN} > 12$, the critical values can be found from the normal approximation to the null distribution of the total number of runs. For instance, using the exact mean and variance of the total number of runs and a continuity correction of 0.5, the left-tail critical region is:

$$\frac{r^i - 0.5 - 2n^i_N n^i_{NN}/n^i}{\sqrt{2n^i_N n^i_{NN}(2n^i_N n^i_{NN} - n^i)/\left[(n^i)^2(n^i-1)\right]}} \leq -z_\alpha \tag{4}$$

Since Swed and Eisenhart (1943) provide the tables with the exact p-values for $n^i_N \leq 20$ and $n^i_{NN} \leq 20$, I use the exact p-values whenever possible.

The null hypothesis is rejected at the α% significance level if either $F(r^i|n^i_N, n^i_{NN}) \leq \alpha/2$ (probability of r^i of less runs is less than or equal to $\alpha/2$) or $1 - F(r^i - 1|n^i_N, n^i_{NN}) \leq \alpha/2$ (probability of r^i of more runs is less than or equal to $\alpha/2$). Table 6 contains the results of the runs tests. There is one rejection at the 5% significance level (Vyacheslav Malafeev, goalkeeper), and two rejections at the 10% level (Bibras Natkho, kicker, in addition to Malafeev).[20] Again, with 41 observations, the expected number of rejections is 2.05 at the 5% significance level and 4.1 at the 10% level. The numbers of resulting rejections are below those proposed by the theory. Therefore, the results of the runs tests suggest that the RFPL players are able to generate random sequences in the field.

Despite the simplicity of the runs test, it has a number of drawbacks. Palacios-Huerta (2014) emphasizes that it has low power to identify a lack of randomness, and can miss many potential sources of dynamic dependence. I test the dynamic dependence by using the method proposed by Arellano and Carrasco (2003). In particular, for each player I estimate the following logit equation:[21]

$$N^i_t = G\left[\alpha_0 + \alpha_1 N^i_{t-1} + \beta_0 N^{i*}_t + \beta_1 N^{i*}_{t-1} + \gamma_1 N^i_{t-1} N^{i*}_{t-1}\right] \tag{5}$$

where $G(x) = exp(x)/\left[1 + exp(x)\right]$, the asterisk * denotes the opponent's choice, and t and $t-1$ mean current and previous choices respectively.

[20] Both players switch their choices not frequently enough.

[21] Palacios-Huerta (2014) uses the specification from Brown and Rosenthal (1990) with two lags. Due to shorter sequences of observations, I use the specification with one lag only.

Table 6: Tests for serial independence of side choices

Player	Nat.	Obs.	Side choice N	Side choice NN	Runs	$1 - F(r-1\|\cdot)$	$F(r\|\cdot)$	AC
Kickers								
Roman Pavlyuchenko	NAT	25	17	8	10	0.866	0.252	0.466
Hulk	FOR	27	14	13	16	0.348	0.788	0.717
Bibras Natkho	FOR	38	21	17	14	0.982	0.039*	0.250
Aleksandr Samedov	NAT	17	8	9	8	0.843	0.319	0.475
Roman Adamov	NAT	16	7	9	8	0.769	0.427	-
Andrey Karyaka	FOR	16	8	8	11	0.214	0.900	0.548
Sergey Kornilenko	FOR	17	10	7	10	0.451	0.743	0.542
Seydou Doumbia	FOR	15	8	7	10	0.296	0.867	0.395
Dmitry Sychev	NAT	15	9	6	8	0.657	0.566	0.921
Aleksandr Kerzhakov	NAT	21	15	6	9	0.735	0.483	0.933
Artem Dzyuba	NAT	24	7	17	14	0.079	0.967	0.890
Dmitry Kombarov	NAT	24	9	15	11	0.783	0.367	0.454
Fedor Smolov	NAT	15	12	3	7	0.270	0.932	0.850
Vágner Love	FOR	21	12	9	12	0.465	0.711	-
Georgi Peev	FOR	21	14	7	12	0.277	0.856	0.468
Goalkeepers								
Sergey Pareiko	FOR	29	16	13	15	0.625	0.521	0.313
Evgeny Gorodov	NAT	26	11	15	13	0.686	0.466	0.896
Ilya Abaev	NAT	23	17	6	12	0.166	0.921	0.733
Anton Shunin	NAT	23	15	8	12	0.490	0.682	-
Sergey Ryzhikov	NAT	52	35	17	27	0.202	0.876	0.157
Artur Nigmatullin	NAT	25	17	8	11	0.748	0.428	-
Soslan Dzhanaev	NAT	30	13	17	14	0.802	0.322	0.422
Mikhail Kerzhakov	NAT	28	18	10	17	0.137	0.943	-
Marinato Guilherme	FOR	49	28	21	24	0.671	0.441	0.136
Anton Kochenkov	FOR	28	20	8	10	0.912	0.175	0.637
Sergey Pesyakov	NAT	36	25	11	19	0.187	0.902	0.505
Aleksandr Selikhov	NAT	28	19	9	12	0.774	0.365	0.586
Stipe Pletikosa	FOR	58	37	21	26	0.745	0.355	0.670
Igor Akinfeev	NAT	75	36	39	37	0.674	0.413	0.883
Andrey Sinitsyn	NAT	20	11	9	9	0.865	0.255	0.684
Andriy Dykan	FOR	27	18	9	13	0.598	0.587	0.492
Yuri Lodygin	NAT	22	14	8	10	0.787	0.369	0.154
David Yurchenko	FOR	35	20	15	19	0.451	0.681	-
Aleksandr Belenov	NAT	51	22	29	29	0.237	0.842	0.001***
Vyacheslav Malafeev	NAT	39	22	17	12	0.998	0.006**	-
Dmitry Khomich	NAT	31	14	17	19	0.214	0.877	-
Vladimir Gabulov	NAT	52	37	15	27	0.077	0.961	0.269
Sergey Narubin	NAT	40	28	12	21	0.150	0.922	0.155
Yuri Zhevnov	FOR	27	19	8	13	0.477	0.726	0.606
Yaroslav Hodzyur	FOR	22	12	10	13	0.395	0.755	0.804
Artem Rebrov	NAT	32	17	15	17	0.561	0.578	0.413

Notes: "NAT" denotes a native player, "FOR" denotes a foreign player.

First, this model generates unbiased and consistent estimates. Second, it allows for unobserved heterogeneity and individual effects to be correlated with the explanatory variables. The results are presented in the right-most column of Table 6. Note that some estimates are missing either because logit equations do not converge or the number of observations is very small. The null hypothesis of randomization, $\alpha_1 = \alpha_2 = \beta_0 = \beta_1 = \gamma_1 = 0$, is rejected for one player only — Aleksandr Belenov (goalkeeper). The rejection is based on the likelihood-ratio test.

4.2 Aggregate Behavior

In this subsection I ask the following question: Can the aggregate behavior of players be considered to be generated from equilibrium play? To answer, I test the hypothesis that the behavior of each player is simultaneously generated by equilibrium play.

Before turning to the aggregate tests for equal success rates and serial independence, I want to check whether the predictions about the aggregate behavior of the kickers and goalkeepers from Chiappori et al. (2002) model hold in my dataset. Table B.2 in Appendix B shows the matrix of penalty kicks' directions. First, we can see that the kickers choose center more frequently than the goalkeepers (263 against 62 kicks). Second, the goalkeepers jump to the kickers' natural side more frequently than the kickers shoot to the natural side (934 against 813 kicks). Third, both the kickers and goalkeepers are more likely to choose the natural side rather than the non-natural side (813 against 603 kicks and 934 against 683 kicks respectively). Fourth, the cell "Natural-Natural" has the greatest number of observations (483 kicks). Thus, the predictions of the model from Chiappori et al. (2002) are supported in my dataset.

4.2.1 Equality of Success Rates

To test the equality of success rates at the aggregate level, I conduct the Pearson joint test and the Kolmogorov-Smirnov test.

The Pearson joint test allows for different success probabilities p^i. For the set of N players, the test statistic is equal to the sum of N individual Pearson statistics and is distributed as chi-squared with N degrees of freedom. The results are reported on the left panel of Table 7. The test statistic for all the players is 34.342 with the corresponding p-value of 0.760. Thus, we do not reject the joint null hypothesis of equal success rates for all the players — $p^i_N = p^i_{NN}$, $\forall i$ — at any reasonable significance level. Recall from Table 5 that there are fewer rejections of the null hypothesis than we might expect in the theory, i.e. when the joint null hypothesis is true. I also conduct the test for different subgroups of players: kickers, goalkeepers, native players, and foreign players. In all the cases the null is not rejected at reasonable significance levels.

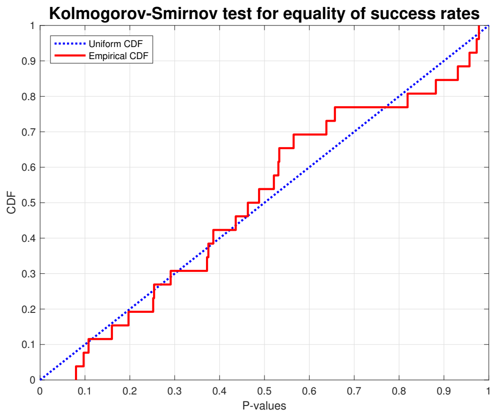
Table 7: Aggregate tests for equality of success rates

		Pearson test		Kolmogorov-Smirnov test	
	N	Statistic	P-value	Statistic	P-value
Kickers	15	12.569	0.636	0.660	0.716
Goalkeepers	26	21.755	0.702	0.649	0.747
All players	41	34.324	0.760	0.804	0.499
Native players	26	18.811	0.844	0.697	0.667
Foreign players	15	15.513	0.415	0.910	0.326
Shootout	16	9.043	0.912	1.158	0.111

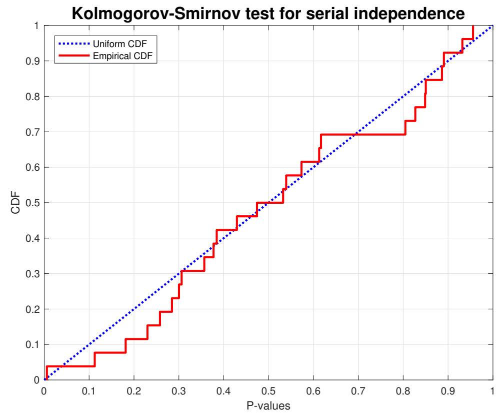
I also make use of the Kolmogorov-Smirnov test that is more powerful than the Pearson joint test against many alternative hypotheses about how the data is generated. The null hypothesis is that an empirical distribution of N observed p-values corresponding to individual Pearson statistics is generated by draws from the uniform distribution $U[0,1]$. Denoting the empirical cumulative distribution function (c.d.f.) of N p-values by $\hat{F}(x)$, the test statistic is $KS = \sqrt{N} \sup_{x\in[0,1]} |\hat{F}(x) - x|$ that has a known distribution.

Figure 1 shows the empirical c.d.f.'s of p-values juxtaposed with the uniform c.d.f. For now we are interested in the left panel. Simple visual comparison suggests that the data may be consistent with the theory predictions as the empirical c.d.f.'s of p-values and the uniform c.d.f. are quite close to each other. The right panel of Table 5 shows that the null hypothesis is not rejected both for all the players and for different subgroups of players at any reasonable significance level.

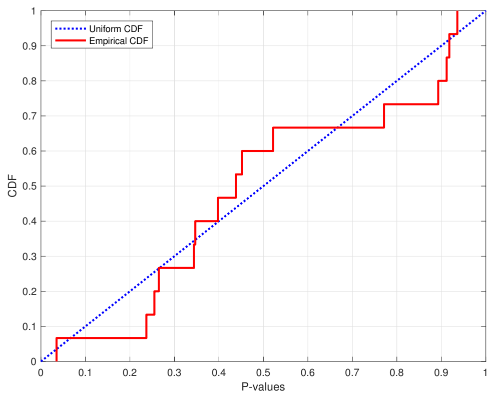
In a nutshell, the data is consistent with the prediction that the success rates are the same across strategies for both the kickers and goalkeepers.



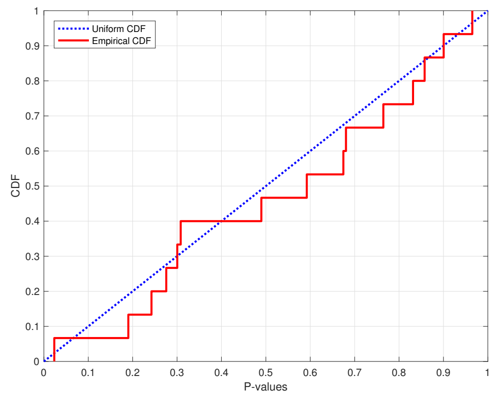
(a) Goalkeepers



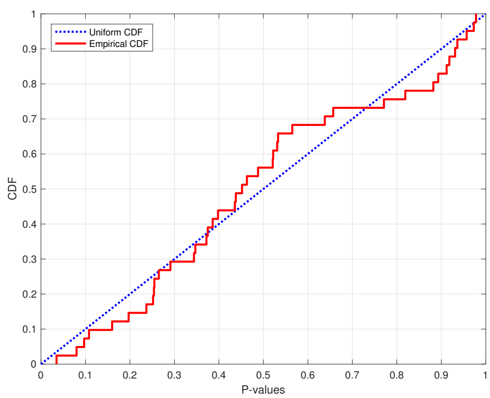
(b) Goalkeepers



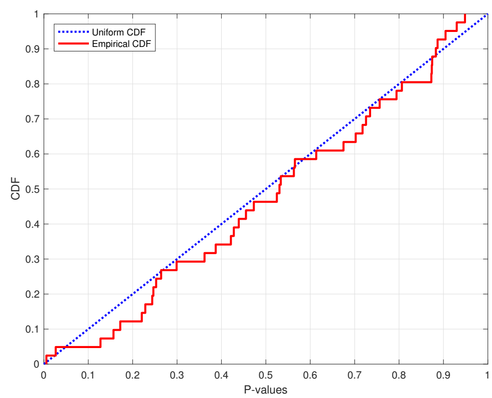
(c) Kickers



(d) Kickers



(e) All players



(f) All players

Figure 1: The blue dotted line depicts the uniform c.d.f. The solid red line depicts the empirical c.d.f. of p-values (left panel) or the empirical c.d.f. of random draw t^i (right panel). Left panel — (a), (c), and (e) — uses the p-values from individual Pearson's chi-square goodness-of-fit tests. Right panel — (b), (d), and (f) — uses a particular random draw of t^i from the runs tests.

4.2.2 Serial Independence of Side Choices

To test the joint hypothesis that the side choices are serially independent for all the players, I conduct the Kolmogorov-Smirnov goodness-of-fit test. For each player i, I draw a random t^i from the uniform distribution $U\left[F(r^i-1|n_N^i,n_{NN}^i),F(r^i|n_N^i,n_{NN}^i)\right]$. Under the null hypothesis, t^i is distributed as $U[0,1]$.

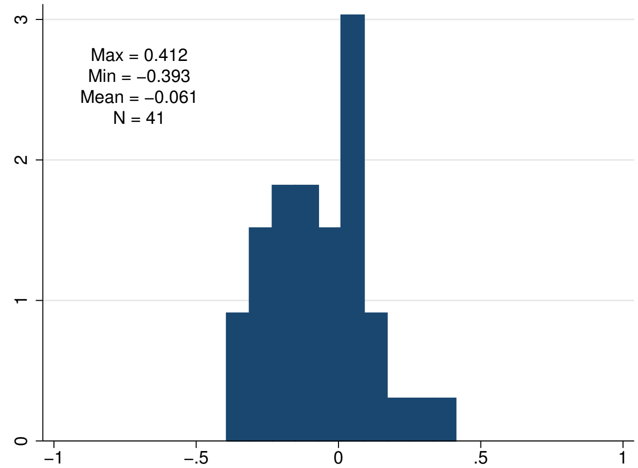
The right panel of Figure 1 depicts the empirical c.d.f.'s of particular realizations of t^i juxtaposed with the uniform c.d.f. Visual comparison suggests that the data may be consistent with the serial independence. I perform 10'000 trials with random draws of t^i, and then calculate the mean Kolmogorov-Smirnov statistic and mean p-value. Table 8 shows the results. We can see that the mean p-value for the set of all the players is 0.513 with a standard deviation of 0.002. The hypothesis of serial independence is not rejected for the kickers, goalkeepers, native and foreign players separately.

Table 8: Aggregate tests for serial independence

	Kolmogorov-Smirnov test	
	KS-statistic	P-value
Kickers	0.641 (0.001)	0.7372 (0.002)
Goalkeepers	0.731 (0.001)	0.611 (0.002)
All players	0.801 (0.001)	0.513 (0.002)
Native players	0.901 (0.001)	0.363 (0.001)
Foreign players	0.825 (0.001)	0.450 (0.001)
Shootout	0.819 (0.001)	0.470 (0.002)

Notes: This table shows the mean Kolmogorov-Smirnov statistics and corresponding p-values over 10'000 trials. Standard deviations are in parentheses.

Figure 2: Serial correlation of side choices



Finally, I estimate the serial correlation of consecutive side choices for all 41 players. Figure 2 shows its empirical density. Serial correlation is statistically significant at the 10% significance level for three players — Artem Dzyuba (-0.393), Bibras Natkho (0.290), and Vyacheslav Malafeev (0.412), and once at the 5% level — Vyacheslav Malafeev. The mean serial correlation coefficient is -0.061. I also calculate the mean serial correlation coefficients separately for the kickers (-0.086) and goalkeepers (-0.046), for the native (-0.052) and foreign (-0.076) players. The differences between the kickers and goalkeepers and between the native and foreign players are not statistically significant (the p-values corresponding to a two-sample t-test are 0.497 and 0.691 respectively).

Summing up, the results of this subsection indicate that both RFPL kickers and goalkeepers' strategic choices are serially independent. This result is in line with the findings of Palacios-Huerta (2003), Palacios-Huerta (2014), and Dohmen and Sonnabend (2018) who also study the behavior of professional athletes in a natural setting.

I show in this section that the kickers and goalkeepers representing RFPL play Minimax. An important difference from the studies listed in the previous paragraph is that they all use the data from the top-5 UEFA leagues, while RFPL is a league that is characterized by lower quality of players. Another study of a low-quality league, Palacios-Huerta (2014) who use the U.S. Major League Soccer[22] data, concludes that players do not play Minimax

[22] There is no apparent observational advantage of the Russian players representing RFPL over the American

there. To the best of my knowledge, my paper is the first one to show that players from a moderate-quality league do play Minimax. They are not as bad as we think.

4.3 Penalty Shootouts

To shed more light on the consistency of the players' behavior with the theoretical predictions, I also consider the sample of penalty kicks performed during penalty shootouts separately. The striking feature of these penalty kicks is that they are performed in rapid succession. For comparison, the lags between two successive penalty kicks performed in two regular games may last for months. On the other hand, recall that the players in a laboratory setting usually play many games in rapid succession. In this sense, penalty shootouts mimic the laboratory experiments. Also recall that the last mostly fail to show that the behavior of people is consistent with the Minimax hypothesis.

The dataset includes 51 penalty shootouts.[23] The teams that shoot first win in 47 percent of cases. This observation somewhat contradicts Apesteguia and Palacios-Huerta (2010) who show that there is a significant advantage for first-movers over second-movers in terms of winning probabilities. On the other hand, Kocher et al. (2012) fail to detect any significant first-mover advantage on a larger sample of penalty shootouts. I restrict the sample to the goalkeepers with at least 10 penalty kicks in penalty shootouts. Table C.1 and Table C.2 in Appendix C show the results. I do not reject neither the hypothesis of equality of success rates, nor the hypothesis of serial independence in strategic choices, i.e. the goalkeepers do play Minimax in penalty shootouts. This finding is in line with Palacios-Huerta (2003), who concludes that the length of time lags does not appear to make any difference for professional soccer players in generating random sequences.

Now I turn to the test proposed by Gauriot et al. (2018) that is more powerful than the tests I use in the current section and that were used in the existing literature on testing the Minimax hypothesis in either a natural or a laboratory settings.

players representing MLS. Palacios-Huerta (2014) notes that even the best MLS players did not play well in Europe. This is comparable with the observation that just four Russian players played in the top-5 leagues between the 2007/08 and 2017/18 seasons.

[23] 48 shootouts in Russian Cup, 3 — in UEFA competitions. Between the 2007/08 and 2017/18 seasons RFPL teams contested in 54 penalty shootouts. So far, I do not have the data on 3 penalty shootouts from Russian Cup 2008.

5 Randomized Fisher Exact Test

Gauriot et al. (2018) develop a randomized test based on the Fisher exact test that rejects the true null hypothesis with exactly probability α at the α significance level. The Kolmogorov-Smirnov test based on this randomized Fisher exact test is valid for any sample size and is more powerful than any existing one used in the literature that test the Minimax hypothesis.[24] As a result, it can potentially overturn the results from the prior literature that make use of less powerful tests. For instance, Gauriot et al. (2018) overturn the results of Hsu et al. (2007). While the last paper concludes that women and junior tennis players do equate the winning probabilities (both the Pearson's joint tests and Kolmogorov-Smirnov tests give the p-values above 0.5), the former one shows that the randomized Fisher exact test places 18.1 percent and 49.2 percent probability weights on the p-values below 0.05 for women and juniors respectively.

In this section I conduct the randomized Fisher exact test, and then I apply the Kolmogorov-Smirnov test to the empirical c.d.f.'s generated from it to reevaluate the predictions about equality of success rates across strategies. The null hypotheses is the same as in Section 4: $p_N^i = p_{NN}^i$ in the individual case and $p_N^i = p_{NN}^i$, $\forall i$, in the joint case. Denote by $f(n_{N,S}^i | n_S^i, n_N^i, n_{NN}^i)$ the probability that the player has $n_{N,S}^i$ successful kicks/jumps after choosing the kicker's natural side, conditional on n_S^i total successful kicks/jumps, after choosing n_N^i and n_{NN}^i kicks/jumps to the natural and non-natural sides respectively. This probability is given by the hypergeometric distribution:

$$f(n_{N,S}^i | n_S^i, n_N^i, n_{NN}^i) = \frac{\binom{n_N^i}{n_{N,S}^i}\binom{n_{NN}^i}{n_S^i - n_{N,S}^i}}{\binom{n_N^i + n_N^i}{n_S^i}} \tag{6}$$

The associated c.d.f. is given by

$$F(n_{N,S}^i | n_S^i, n_N^i, n_{NN}^i) = \sum_{j=\max(n_S^i - n_{NN}^i, 0)}^{n_N^i} f(j | n_S^i, n_N^i, n_{NN}^i) \tag{7}$$

Table 9 shows the c.d.f.'s that I use to perform the randomized Fisher exact test.

[24] See Gauriot et al. (2018) for the details.

Table 9: C.d.f.'s for the Fisher exact test

Player	Nat.	Obs.	n_N	n_{NN}	n_S	$F(n_{N,S}-1\|\cdot)$	$F(n_{N,S}\|\cdot)$
Kickers							
Roman Pavlyuchenko	NAT	25	17	19	10	0.349	0.726
Hulk	FOR	27	14	13	21	0.362	0.714
Bibras Natkho	FOR	38	21	17	31	0.302	0.624
Aleksandr Samedov	NAT	17	8	9	13	0.241	0.665
Roman Adamov	NAT	16	7	9	12	0.192	0.608
Andrey Karyaka	FOR	16	8	8	13	0.500	0.900
Sergey Kornilenko	FOR	17	10	7	13	0.088	0.441
Seydou Doumbia	FOR	15	8	7	12	0.554	0.923
Dmitry Sychev	NAT	15	9	6	14	0.000	0.600
Aleksandr Kerzhakov	NAT	21	15	6	19	0.000	0.500
Artem Dzyuba	NAT	24	7	17	17	0.682	0.923
Dmitry Kombarov	NAT	24	9	15	21	0.042	0.308
Fedor Smolov	NAT	15	12	3	13	0.629	0.971
Vágner Love	FOR	21	12	9	16	0.039	0.258
Georgi Peev	FOR	21	14	7	19	0.900	1.000
Goalkeepers							
Sergey Pareiko	FOR	29	16	13	15	0.353	0.664
Evgeny Gorodov	NAT	26	11	15	13	0.345	0.687
Ilya Abaev	NAT	23	17	6	12	0.271	0.731
Anton Shunin	NAT	23	15	8	12	0.334	0.712
Sergey Ryzhikov	NAT	52	35	17	27	0.426	0.675
Artur Nigmatullin	NAT	25	17	8	11	0.393	0.754
Soslan Dzhanaev	NAT	30	13	17	14	0.515	0.805
Mikhail Kerzhakov	NAT	28	18	10	17	0.519	0.810
Marinato Guilherme	FOR	49	28	21	24	0.181	0.403
Anton Kochenkov	FOR	28	20	8	10	0.123	0.448
Sergey Pesyakov	NAT	36	25	11	19	0.571	0.852
Aleksandr Selikhov	NAT	28	19	9	12	0.527	0.882
Stipe Pletikosa	FOR	58	37	21	26	0.649	0.840
Igor Akinfeev	NAT	75	36	39	37	0.686	0.837
Andrey Sinitsyn	NAT	20	11	9	9	0.098	0.396
Andriy Dykan	FOR	27	18	9	13	0.661	0.906
Yuri Lodygin	NAT	22	14	8	10	0.670	0.909
David Yurchenko	FOR	35	20	15	19	0.100	0.306
Aleksandr Belenov	NAT	51	22	29	29	0.777	0.911
Vyacheslav Malafeev	NAT	39	22	17	12	0.787	0.933
Dmitry Khomich	NAT	31	14	17	19	0.776	0.937
Vladimir Gabulov	NAT	52	37	15	27	0.809	0.966
Sergey Narubin	NAT	40	28	12	21	0.838	0.975
Yuri Zhevnov	FOR	27	19	8	13	0.856	1.000
Yaroslav Hodzyur	FOR	22	12	10	13	0.882	0.988
Artem Rebrov	NAT	32	17	15	17	0.913	0.986

Notes: "NAT" denotes a native player, "FOR" denotes a foreign player.

I randomly draw a test statistic f^i from the uniform distribution $U\left[0, F(n^i_{N,S}|n^i_S, n^i_N, n^i_{NN})\right]$ if $n^i_{N,S} = n^i_S - n^i_{NN}$, and from $U\left[F(n^i_{N,S} - 1|n^i_S, n^i_N, n^i_{NN}), F(n^i_{N,S}|n^i_S, n^i_N, n^i_{NN})\right]$ otherwise. Under the null hypothesis, f^i is distributed as $U[0,1]$. For each player i, the null hypothesis is rejected at the 5% significance level if either $f^i \leq 0.025$ or $f^i \geq 0.975$.

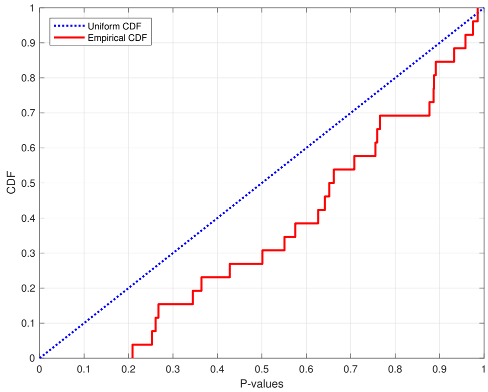
Since f^i is a random statistic, I perform 10'000 trials for each player and then calculate the mean frequency at which the null hypothesis is rejected. The resulting rejection rate for 41 players at the 5% significance level is 1.92 percent, while at the 10% significance level — 5.62 percent. Both rejection rates are below ones predicted by the theory. Therefore, at the individual level the rates at which equality of success rates is rejected at the reasonable significance levels are consistent with the theory.

To test the joint hypothesis that success rates are equal for each player, I apply the Kolmogorov-Smirnov test to the empirical c.d.f.'s generated from the randomized Fisher exact test. The procedure follows one I describe in Section 4. The results are shown in Figure 3, Figure 4, and Table 10.

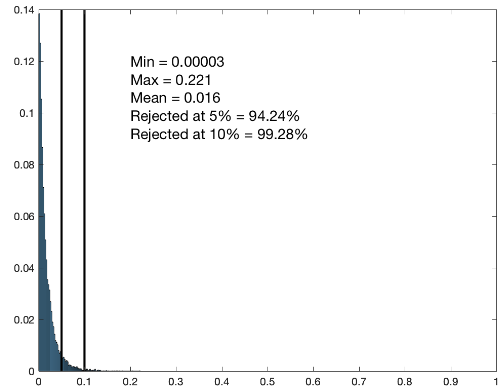
Table 10: Kolmogorov-Smirnov test based on the randomized Fisher exact test

		P-value			Rejections	
	KS-statistic	Min	Max	Mean	5% level	10% level
Kickers	0.627 (0.001)	0.105	1.000	0.749 (0.002)	0%	0%
Goalkeepers	1.591 (0.002)	0.000	0.221	0.016 (0.000)	94.24%	99.28%
All players	1.335 (0.002)	0.000	0.482	0.069 (0.001)	49.09%	77.65%
Native players	1.383 (0.002)	0.000	0.540	0.055 (0.001)	60.69%	83.95%
Foreign players	0.726 (0.001)	0.097	0.9995	0.608 (0.002)	0%	0.01%

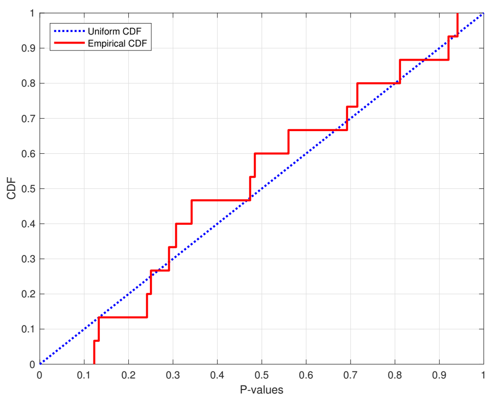
Notes: The null hypothesis is $p^i_N = p^i_{NN}$, $\forall i$. Columns 1 and 4 show the mean Kolmogorov-Smirnov statistics and corresponding p-values over 10'000 trials. Standard deviations are in parentheses.



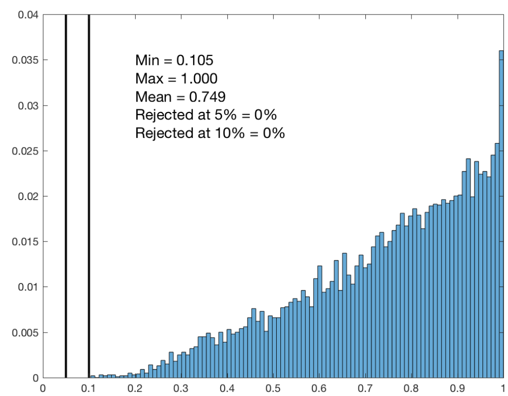
(a) Goalkeepers



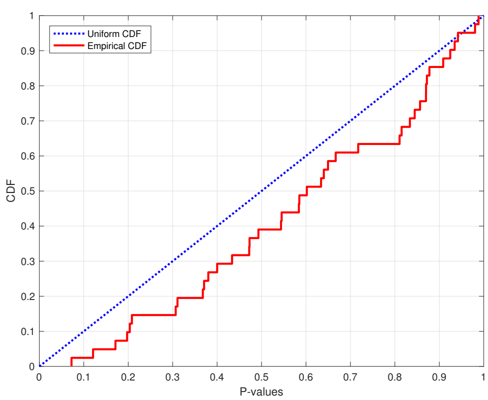
(b) Density of KS test p-values



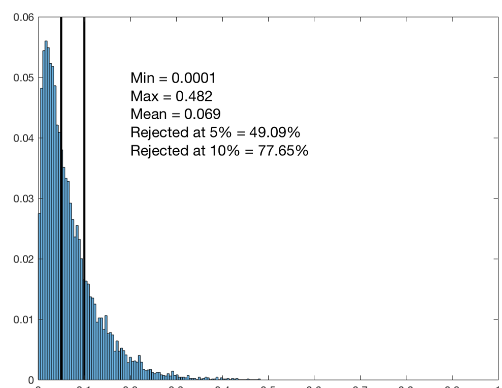
(c) Kickers



(d) Density of KS test p-values

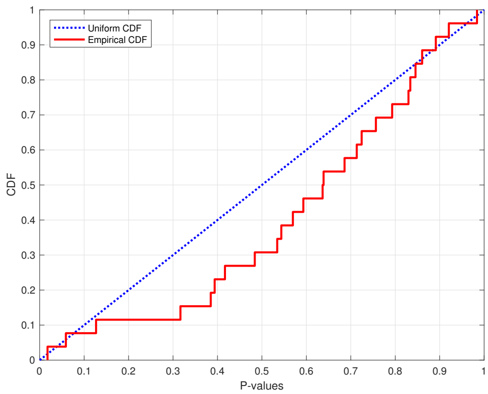


(e) All players

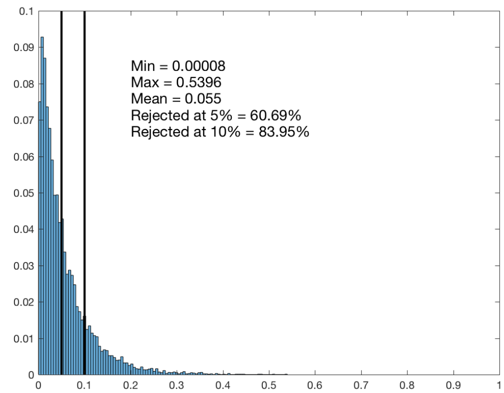


(f) Density of KS test p-values

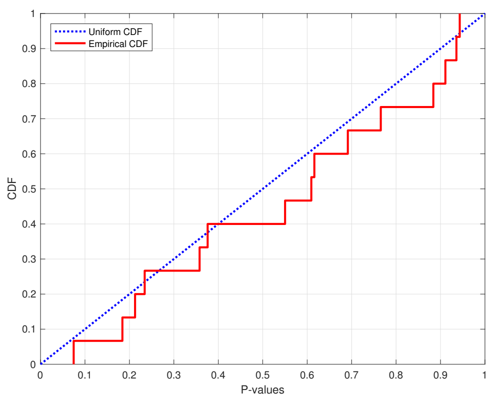
Figure 3: The blue dotted line on the left panel depicts the uniform c.d.f. The solid red line on the left panel depicts the empirical c.d.f. of a particular random draw f^i from the randomized Fisher exact test. Vertical black solid lines on the right panel correspond to the 5% and 10% significance levels. Vertical bars show the density of the p-values resulted from 10'000 trials of the Kolmogorov-Smirnov test based on the randomized Fisher exact test.



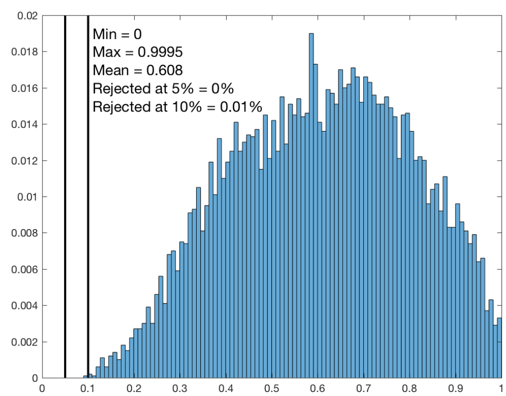
(a) Native players



(b) Density of KS test p-values



(c) Foreign players



(d) Density of KS test p-values

Figure 4: The blue dotted line on the left panel depicts the uniform c.d.f. The solid red line on the left panel depicts the empirical c.d.f. of a particular random draw f^i from the randomized Fisher exact test. Vertical black solid lines on the right panel correspond to the 5% and 10% significance levels. Vertical bars show the density of the p-values resulted from 10’000 trials of the Kolmogorov-Smirnov test based on the randomized Fisher exact test.

We can see that some results from Section 4 are overturned. First, the joint null hypothesis of equal success rates is rejected at the 5% significance level for almost half of trials (49.09 percent), and for about three-fourths at the 10% level (77.65 percent). In other words, if we run the test once, it is almost equally likely to reject the null hypothesis at the 5% significance level, and three out of four times at the 10% level. Second, the joint null is rejected in 60.69 percent of trials at the 5% level (83.95 percent — at the 10% level) for the native players. This result is consistent with the idea that the players of lower quality do not have enough Minimax strategic skills as proposed by Palacios-Huerta (2014). However, note that we come to this conclusion after using the Kolmogorov-Smirnov test based on the randomized Fisher exact test rather than the tests from Section 4. The subgroup consisting of all the goalkeepers experience the most striking failure: at the 5% level the joint null is rejected in more than 94 percent of trials.

However, there exist subgroups for which the joint null hypothesis of equality of success rates does not even come close to being rejected. The first one is the subgroup consisting of all the kickers: no single rejections at either 5% or 10% significance levels. The second one is the subgroup of foreign players: no rejections at the 5% level and only one rejection at the 10% level.

Further calculations reveal that the main driver of the failures is the behavior of native goalkeepers. For this subgroup, the joint null hypothesis is rejected in 98.19 percent of trials at the 5% significance level with the mean p-value of 0.009. In turn, there is no single rejection of the joint null even at the 10% significance level for the subgroups of native kickers or foreign goalkeepers. Furthermore, when I perform the Kolmogorov-Smirnov test on the sample without native goalkeepers, I get that there are no rejections of the joint null at the 5% level and just six rejections at the 10% level. Thus, without the subgroup of native goalkeepers, the behavior of the players representing RFPL is consistent with the Minimax hypothesis.

Again, it is not as bad as we think.

6 Conclusion

This paper uses a unique dataset of penalty kicks performed by the soccer players representing RFPL to test whether their behaviour in a natural setting is consistent with the predictions of the Minimax hypothesis. While there exist several published papers that perform a similar exercise, I come to some interesting and challenging conclusions.

The contribution of the paper is threefold. First, I construct the dataset comprised of 1679 penalty kicks which is a novelty by itself. This dataset covers literally *all* the penalty kicks with RFPL players on one or another side (or both) of the penalty mark. Second, I show that the behavior of players representing RFPL, a league of *lower quality* than the top-5 UEFA leagues, is consistent with the Nash equilibrium predictions. For this purpose

I employ the tests that were used in the prior papers on Minimax behavior. In fact, I conclude that (i) success rates are the same across strategies for all the players, and (ii) both the kickers and goalkeepers' strategic choices are serially independent. These results hold when I consider the different subgroups of players — kickers, goalkeepers, native players, and foreign players — as well. To the best of my knowledge, this is the first paper to show the consistency with the Minimax predictions for players representing a moderate-quality league. The previous attempt by Palacios-Huerta (2014) — on the data from the U.S. Major League Soccer — shows that the players do not play Minimax there. This suggests that even in moderate-quality leagues players can have Minimax strategic skills. Finally, I make use of the test developed by Gauriot et al. (2018) that is more powerful than any existing one used in the literature that test the Minimax hypothesis. I believe, this paper is the first one that performs this test on the penalty kicks data. I show that on the aggregate level the result that success rates are the same across strategies for all the players has to be overturned. In addition, it has to be overturned for the subgroups of goalkeepers and native players as well. However, it still holds for strikers and foreign players. I show that the main driver of the failures is the behavior of native goalkeepers.

Since the papers that have used the data on penalty kicks are those that mostly confirmed the predictions of the Minimax hypothesis, my conclusions are like a wake-up call that these positive results should be taken with a grain of salt.

References

Apesteguia, J. and I. Palacios-Huerta (2010): "Psychological Pressure in Competitive Environments: Evidence from a Randomized Natural Experiment," *American Economic Review*, 100(5), 2548–2564.

Arellano, M. and R. Carrasco (2003): "Binary Choice Panel Data Models with Predetermined Variables," *Journal of Econometrics*, 115(1), 125–157.

Azar, O. H. and M. Bar-Eli (2011): "Do Soccer Players Play the Mixed-Strategy Nash Equilibrium?" *Applied Economics*, 43(25), 3591–3601.

Bar-Eli, M., O. H. Azar, I. Ritov, Y. Keidar-Levin, and G. Schein (2007): "Action Bias Among Elite Soccer Goalkeepers: The Case of Penalty Kicks," *Journal of Economic Psychology*, 28(5), 606–621.

Baumann, F., T. Friehe, and M. Wedow (2011): "General Ability and Specialization: Evidence From Penalty Kicks in Soccer," *Journal of Sports Economics*, 12(1), 81–105.

Brown, J. N. and R. W. Rosenthal (1990): "Testing the Minimax Hypothesis: A Re-Examination of O'Neill's Game Experiment," *Econometrica*, 58(5), 1065–1081.

Cain, L. P. and D. D. Haddock (2006): "Measuring Parity: Tying into the Idealized Standard Deviation," *Journal of Sports Economics*, 7(3), 330–338.

Chiappori, P.-A., S. Levitt, and T. Groseclose (2002): "Testing Mixed-Strategy Equilibria When Players Are Heterogeneous: The Case of Penalty Kicks in Soccer," *American Economic Review*, 92(4), 1138–1151.

Coloma, G. (2007): "Penalty Kicks in Soccer: An Alternative Methodology for Testing Mixed-Strategy Equilibria," *Journal of Sports Economics*, 8(5), 530–545.

Dohmen, T. and H. Sonnabend (2018): "Further Field Evidence for Minimax Play," *Journal of Sports Economics*, 19(3), 371–388.

FIFA (2017): *Laws of the Game 2017/18*, Fédération Internationale de Football Association.

Fort, R. (2007): "Comments on "Measuring Parity"," *Journal of Sports Economics*, 8(6), 642–651.

Gauriot, R., L. Page, and J. Wooders (2018): "Nash at Wimbledon: Evidence from Half a Million Serves," *Mimeo*.

Gibbons, J. D. and S. Chakraborti (2003): *Nonparametric Statistical Inference (Fourth Ed.)*, Marcel Dekker, Inc.

Hsu, S.-H., C.-Y. Huang, and C.-T. Tang (2007): "Minimax Play at Wimbledon: Comment," *American Economic Review*, 97(1), 517–523.

Kocher, M. G., M. V. Lenz, and M. Sutter (2012): "Psychological Pressure in Competitive Environments: New Evidence from Randomized Natural Experiments," *Management Science*, 58(8), 1585–1591.

Kovash, K. and S. D. Levitt (2009): "Professionals Do Not Play Minimax: Evidence from Major League Baseball and the National Football League," NBER working paper 15347, National Bureau of Economic Research.

Levitt, S. D., J. A. List, and D. H. Reiley (2010): "What Happens in the Field Stays in the Field: Exploring Whether Professionals Play Minimax in Laboratory Experiments," *Econometrica*, 78(4), 1413–1434.

Misirlisoy, E. and P. Haggard (2014): "Asymmetric Predictability and Cognitive Competition in Football Penalty Shootouts," *Current Biology*, 24(16), 1918–1922.

Moschini, G. (2004): "Nash Equilibrium in Strictly Competitive Games: Live Play in Soccer," *Economics Letters*, 85(3), 365–371.

Noll, R. G. (1988): "Professional Basketball," *Studies in Industrial Economics Paper No. 144*.

O'Neill, B. (1987): "Nonmetric Test of the Minimax Theory of Two-Person Zerosum Games," *Proceedings of the National Academy of Sciences*, 84(7), 2106–2109.

Owen, P. D. (2012): "Measuring Parity in Sports Leagues with Draws: Further Comments," *Journal of Sports Economics*, 13(1), 85–95.

Palacios-Huerta, I. (2003): "Professionals Play Minimax," *Review of Economic Studies*, 70(2), 395–415.

——— (2014): *Beautiful Game Theory: How Soccer Can Help Economics*, Princeton University Press.

Palacios-Huerta, I. and O. Volij (2008): "Experientia Docet: Professionals Play Minimax in Laboratory Experiments," *Econometrica*, 76(1), 71–115.

Palomino, F., L. Rigotti, and A. Rustichini (2000): "Skill, Strategy and Passion: An Empirical Analysis of Soccer," *Mimeo.*

Quirk, J. and R. D. Fort (1997): *Pay Dirt: The Business of Professional Team Sports*, Princeton University Press.

Scully, G. W. (1989): "The Business of Major League Baseball," *University of Chicago Press.*

Swed, F. S. and C. Eisenhart (1943): "Tables for Testing Randomness of Grouping in a Sequence of Alternatives," *The Annals of Mathematical Statistics*, 14(1), 66–87.

Walker, M. and J. Wooders (2001): "Minimax Play at Wimbledon," *American Economic Review*, 91(5), 1521–1538.

Wooders, J. (2010): "Does Experience Teach? Professionals and Minimax Play in the Lab," *Econometrica*, 78(3), 1143–1154.

Appendix

A Competitive Balance

Since the comprehensive discussion of competitive balance issues is beyond the scope of this paper, I restrict myself to using one widely used measure — *relative standard deviation* (henceforth RSD). This approach was pioneered by Noll (1988) and Scully (1989).[25] Originally, this measure did not take into account the possibility of draws that are quite common in soccer. The approach was developed under the assumption of 50-50 winning probabilities in an ex-ante perfectly competitive balanced league. Subsequent work, including the discussion between Cain and Haddock (2006) and Fort (2007), and later Owen (2012), shed light on this issue. In particular, Cain and Haddock (2006) using the data from English Premier League between the 1888/89 and 2003/04 seasons show that the frequency of draws is 24.59 percent. Thus, assigning a zero weight on draws is a very restrictive assumption.

The idea behind RSD allowing for draws is the following. First, calculate the standard deviation of the share of total points earned *within a season* (points earned within a season/total possible points within a season). In the literature this is referred as the actual standard deviation (henceforth ASD). Note that I use the percentage of points rather than absolute points. To measure the competitive balance, ASD is compared with the so-called idealized standard deviation (henceforth ISD) corresponding to an ex ante perfectly balanced league where all the teams are of equal strength. In particular,

$$RSD = \frac{ASD}{ISD} \tag{A.1}$$

To derive ISD, let q be the probability of a draw. Denote the expected value and the standard deviation of the share of total season points earned in *one game* (points earned in a given game/total possible points within a season) by μ_p and σ_P respectively. Consider a league where each team plays N games per season.

[25] See Quirk and Fort (1997) for the discussion.

Under the 3-1-0 system (win — 3 points, draw — 1 point, loss — 0 points), their values are given by[26]

$$\mu_P = \frac{0.5(3-q)}{3N} \tag{A.2}$$

$$\sigma_P = \sqrt{\frac{(1-q)(q+9)}{36N}} \tag{A.3}$$

Since I calculated these two measures under the assumption that ex ante all the teams are of equal strength, the standard deviation σ_P is actually ISD.

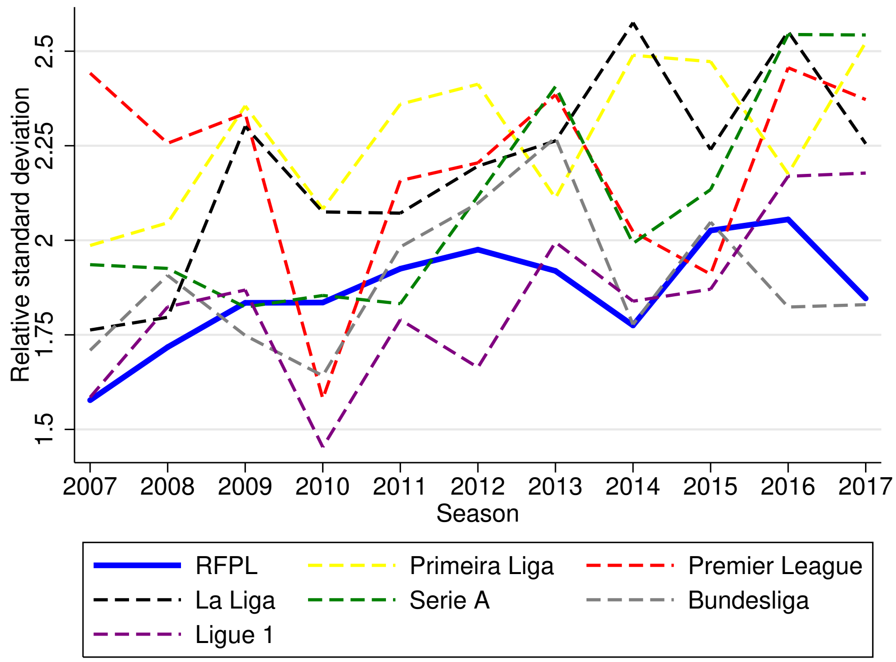
I measure the competitive balance of the top-7 UEFA soccer leagues — Premier League, La Liga, Serie A, Bundesliga, Ligue 1, RFPL, and Primeira Liga — for the 2007/08-2017/18 seasons. First, I calculate ASD — the standard deviation of the share of total points earned for each league across all the seasons. Second, I calculate the empirical analogue of q — frequency of draws across leagues and seasons — that equals to 25.7 percent. This fraction is pretty close to the estimate from Cain and Haddock (2006). Thus, the ISD estimate is

$$\hat{\sigma}_P = \sqrt{\frac{0.191}{N}} \tag{A.4}$$

Dynamics of resulting RSD's across leagues is shown on Figure A.1. The most competitive league, as measured by the average RSD, is Ligue 1 (average RSD — 1.84), while the least competitive — Primeira Liga (2.274). RFPL is ranked the second most competitive league with the average RSD of 1.862.

[26] The expected share of total points earned in one game is given by $0.5(1-q)\cdot 1/N + q \cdot 1/3N$. The expression for the standard deviation is derived straightforwardly by definition. See Owen (2012) for the discussion.

Figure A.1: Competitive balance in the top-7 soccer leagues, 2007/08-2017/18 seasons



Notes: RFPL — Russia (average relative standard deviation over 2007-2018 is 1.862), Primeira Liga — Portugal (2.274), Premier League — England (2.193), La Liga — Spain (2.190), Serie A — Italy (2.101), Bundesliga — Germany (1.894), Ligue 1 — France (1.840). X-axis indicates the beginning of the season (e.g., 2017 means the 2017/18 season). For the 2011/12 season in RFPL I use only the first 30 games of all the teams (excluding the second part of the season when the teams ranked 1-8 played against each other and the teams ranked 9-16 played against each other). Primeira Liga was extended from 16 to 18 teams in the 2014/15 season.

B Additional Tables

Table B.1: Simultaneity of the kicker and goalkeeper's actions

	Dependent variable: Kicker chooses the natural side			
	(1)	(2)	(3)	(4)
Goalkeeper chooses the natural side	0.0458 (0.0294)	0.0115 (0.0377)	0.0451 (0.0296)	0.0108 (0.0382)
Fraction of natural-side kicks excluding given penalty kick	0.3806*** (0.0475)	0.3703*** (0.0913)	0.3715*** (0.0477)	0.3656*** (0.0918)
Fraction of natural-side jumps excluding given penalty kick	0.0826 (0.0813)	0.0539 (0.0987)	0.0832 (0.0821)	0.0629 (0.0995)
Year × competition dummies	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes
Kickers with at least 5 kicks	No	Yes	No	Yes
Adj. R^2	0.0436	0.0120	0.0432	0.0123
Number of observations	1181	759	1181	759

Notes: *** $p < 0.01$. *Standard errors are in parentheses.*

Table B.2: Observed matrix of penalty kicks' directions

	Goalkeeper			
Kicker	Natural	Center	Non-natural	Total
Natural	483	17	313	813
Center	136	31	96	263
Non-natural	315	14	274	603
Total	934	62	683	1679

C Penalty Shootouts

Table C.1: Tests for equality of success rates in penalty shootouts

			Side choice		Success rate			
Player	Nat.	Obs.	N	NN	N	NN	Pearson	P-value
Yuri Lodygin	NAT	10	80.0	20.0	50.0	50.0	0.000	1.000
Vyacheslav Malafeev	NAT	10	80.0	20.0	50.0	50.0	0.000	1.000
Vladimir Gabulov	NAT	17	70.6	29.4	16.7	20.0	0.027	0.870
Aleksandr Selikhov	NAT	16	62.5	37.5	20.0	16.7	0.027	0.869
Sergey Pareiko	FOR	11	54.5	45.5	50.0	60.0	0.110	0.740
Igor Akinfeev	NAT	21	42.9	57.1	33.3	41.7	0.151	0.697
Marinato Guilherme	FOR	19	57.9	42.1	27.3	37.5	0.224	0.636
Stipe Pletikosa	FOR	27	63.0	37.0	29.4	40.0	0.318	0.573
Artur Nigmatullin	NAT	21	71.4	28.6	20.0	33.3	0.420	0.517
Sergey Pesyakov	NAT	19	73.7	26.3	35.7	20.0	0.421	0.516
David Yurchenko	FOR	10	50.0	50.0	20.0	40.0	0.476	0.490
Sergey Ryzhikov	NAT	14	64.3	35.7	22.2	40.0	0.498	0.480
Aleksandr Belenov	NAT	11	36.4	67.6	50.0	28.6	0.505	0.477
Dmitry Khomich	NAT	15	40.0	60.0	50.0	22.2	1.250	0.264
Anton Kochenkov	FOR	11	81.8	18.2	11.1	50.0	1.664	0.197
Artem Rebrov	NAT	17	70.6	29.4	41.7	0.0	2.951	0.086*

Table C.2: Tests for serial independence of side choices in penalty shootouts

			Side choice				
Player	Nat.	Obs.	N	NN	Runs	$1 - F(r-1\|\cdot)$	$F(r\|\cdot)$
Yuri Lodygin	NAT	10	8	2	4	0.778	0.533
Vyacheslav Malafeev	NAT	10	8	2	4	0.778	0.533
Vladimir Gabulov	NAT	17	12	5	10	0.181	0.925
Aleksandr Selikhov	NAT	16	10	6	7	0.863	0.287
Sergey Pareiko	FOR	11	6	5	6	0.738	0.522
Igor Akinfeev	NAT	21	9	12	10	0.795	0.362
Marinato Guilherme	FOR	19	11	8	10	0.648	0.547
Stipe Pletikosa	FOR	27	17	10	17	0.112	0.955
Artur Nigmatullin	NAT	21	15	6	10	0.517	0.668
Sergey Pesyakov	NAT	19	14	5	11	0.093	0.740
David Yurchenko	FOR	10	5	5	8	0.167	0.960
Sergey Ryzhikov	NAT	14	9	5	7	0.713	0.510
Aleksandr Belenov	NAT	11	4	7	5	0.858	0.333
Dmitry Khomich	NAT	15	6	9	7	0.825	0.343
Anton Kochenkov	FOR	11	9	2	4	0.800	0.491
Artem Rebrov	NAT	17	12	5	8	0.635	0.579
Sergey Narubin	NAT	12	8	4	8	0.212	0.929