Minimax on the gridiron: Serial correlation and its effects on outcomes in the National Football League

Noha Emara and David Owens and John Smith and Lisa Wilmer

Rutgers University-Camden, Haverford College

26. September 2014

Minimax on the gridiron: Serial correlation and its effects on outcomes in the National Football League*

Noha Emara[†] David Owens[‡] John Smith[§] Lisa Wilmer[¶]

September 26, 2014

Abstract

We examine whether the predictions of minimax in zero-sum games holds under highly incentivized conditions with highly informed informed decision makers. We examine data from 3455 National Football League (NFL) games from the 2000 season through the 2012 season. We categorize every relevant play as either a rush or a pass. We find that, despite the predictions of minimax, the pass-rush mix exhibits negative serial correlation. In other words, given the conditions of the play, teams employ an exploitable strategy in that play types alternate more frequently than implied by an independent stochastic process. We also find that the efficacy of plays are affected by previous actions and previous outcomes in a manner that is not consistent with minimax. Our analysis suggests that teams could profit from more clustered play selections, which switch play type less frequently. Our results are consistent with the explanation that teams excessively switch play types in order to not be perceived as predictable.

Keywords: serial correlation, game theory, mixed strategies, matching pennies

JEL: C72, C93, D03

*The authors thank Lisa Saal for helpful technical assistance with the large data set. The authors thank Roberto Barbera, Sean Duffy, Matthew Jones, and participants at the Second Haverford Conference on Behavioral and Experimental Economics for helpful comments. This research is supported by Rutgers Research Council Grant #202071.

†Rutgers University-Camden, Department of Economics

‡Haverford College, Department of Economics

§Corresponding Author; Rutgers University-Camden, Department of Economics, 311 North 5th Street, Camden, New Jersey, USA 08102; Email: smithj@camden.rutgers.edu; Phone: +1 856 225-6319; Fax: +1 856 225-6602.

¶Rutgers University-Camden, Department of Economics

1 Introduction

In two-player, zero-sum games with a unique Nash equilibrium in mixed strategies, it is incumbent on the players to mix according to minimax. In particular, the mixing cannot be predictable, otherwise a player could devise a strategy to exploit an opponent who does not properly mix.

A natural question is whether observed mixed strategies occur as predicted by minimax. Although it is well-known that people have difficulty detecting and producing random sequences of the sort required for the execution of minimax,[1] research shows that there are differences between the generation of such sequences in decision problems and strategic settings.[2] However, laboratory evidence suggests that mixing often does not occur as predicted, particularly when the data is analyzed at the individual level.[3]

On the other hand, these studies are vulnerable to the critique that subjects do not have sufficient material incentives to employ a strategy that cannot be exploited and they do not have sufficient incentives to detect an exploitable strategy of their opponent. Further, due to their unfamiliarity with the strategic issues, the subjects might lack the experience to adequately play the game.

Rather than investigate this question in the laboratory, we go to the field, literally. We examine strategic decisions in the National Football League (NFL). We observe the type of play called by the teams and investigate whether the sequence of plays conforms to the predictions of minimax. While we acknowledge that the laboratory has certain advantages over the field, it is also the case that our setting exhibits advantages over the laboratory. NFL head coaches are paid large salaries and are under intense pressure to win, as evidenced by their frequent employment terminations. They also can make their decisions through

[1] For instance, see Wagenaar (1972), Bar-Hillel, and Wagenaar (1991), Rabin (2002), and Oskarsson, Van Boven, McClelland, and Hastie (2009).

[2] See Rapoport and Budescu (1992) and Budescu and Rapoport (1994).

[3] Since O'Neill (1987) and the reexamination of the original data by Brown and Rosenthal (1990) there has been mixed evidence regarding mixed strategies in the laboratory. This literature includes Batzilis et al. (2013), Binmore, Swierzbinski, and Proulx (2001), Geng, Peng, Shachat, and Zhong (2014), Mookherjee and Sopher (1994, 1997), O'Neill (1991), Ochs (1995), Palacios-Huerta and Volij (2008), Rapoport and Amaldoss (2000, 2004), Rapoport and Boebel (1992), Rosenthal, Shachat, and Walker (2003), Shachat (2002), Van Essen and Wooders (2013).

conferring with other members of the coaching staff.[4] Further, their decisions are not made without sufficient deliberation, as each team makes detailed plans specific to the game.

We find that, despite these incentives for success, the level of expertise, the ability to consult others, and the ability to make detailed plans prior to the game, play calling exhibits negative serial correlation. We also find that, according to two measures, the efficacy of plays are affected by previous actions and previous outcomes. In other words, teams excessively switch play types and this has negative consequences for the efficacy of the plays. Our results suggest that teams could profit from more clustered play sequences, which switch play type less frequently.

While the literature, which finds that people have difficulty producing statistically independent sequences, can explain the negative serial correlation, it seems to not provide a satisfactory explanation for the reduced efficacy of plays associated with the negative serial correlation. In fact, our results are consistent with the explanation that the defenses expect the negative serial correlation exhibited by the offenses. This seems to be the case because clustered plays have a larger efficacy. This leads to the question of why the offense would elect to employ an exploitable strategy that exhibits negative serial correlation. It is possible that the offenses excessively switch in order to appear "unpredictable" to people (fans, owners, etc.) who have trouble detecting statistically independent sequences. It could also be the case that this effect is larger than the negative consequences that arise from the negative serial correlation. Therefore, our analysis is consistent with the view that teams want to be viewed as unpredictable and that they accept the reduced efficacy of their plays resulting from the negative serial correlation.

1.1 Background details of football

American football (hereafter referred to as *football*) is contested on a 100 yard[5] long rectangular field. Two competing teams of eleven players attempt to advance a ball towards the other's *end zone*, located at opposite ends of the field. Teams receive six points from a *touchdown*,

[4] Okano (2013) finds that behavior in a repeated game with a unique mixed strategy equilibrium is closer to the minimax prediction when teams of two play rather than when individuals play.

[5] One yard is the equivalent of 0.9144 meters.

by advancing the ball into their opponent's end zone, and three points from a *field goal*, by kicking the ball through a set of elevated goal posts over their opponent's end zone.

Football is unique among sports in two respects. First, there is a clearer distinction between offense and defense than in other sports. There is a stoppage in play each time possession of the ball transfers from one team to the other, and teams nearly always replace all eleven players when this occurs. Second, the action is broken into discrete units called *plays*. The offensive team increases its chance of scoring points, and therefore winning the game, by advancing the ball towards the defensive team's end zone. The distance to the defensive team's end zone is referred to as *field position*.

When a team has possession of the ball, it has four plays, called *first down* through *fourth down*, to advance the ball a total of ten yards. These plays are referred to as plays from *scrimmage*. If the team succeeds in advancing ten yards, an achievement also referred to as a *first down*, the offensive team gets another set of four downs to advance the ball another ten yards. If they fail to net ten yards in a set of four plays, the ball is awarded to the other team.[6] The number of yards that the offensive team needs to advance in order to achieve a first down is called the *distance*. The sequence of plays where only one team possess the ball is referred to as a *possession*.

From an analytic perspective, an attractive aspect of football is that the beginning of each play can be well-characterized by the score, down, distance, and field position. There is an extremely large number of strategies that teams can employ for a specified play. For the offense, those plays fall into two categories: a *pass* or a *rush*. A pass is a play in which one player[7] attempts to throw the ball forward to another player, who attempts to catch it. A rush is a play in which a player attempts to advance the ball by carrying it.

It is important that the offense and defense are not predictable by the opposition. If the defense knew the play called by the offense, they could devise a strategy to best defend against the play. If the offense knew the strategy of the defense, they could devise a play

[6] The offensive team has the option to "punt," or kick the ball down the field, surrendering the ball to the other team. Teams often employ the punt on fourth down, making third down effectively the final opportunity to complete the ten yards. See Romer (2006) for more on the decision to punt on fourth down.

[7] Nearly always, this passing player is the quarterback.

to best attack the defense. In short, the play called by the offense and the strategy of the defense needs to selected in an unpredictable manner. While we do not observe the strategy employed by the defense, we do observe whether the play of the offense is a pass or a rush. We are interested in whether the pass-rush mix is consistent with the predictions of minimax.[8]

To facilitate an understanding of the following analysis, we describe a few additional details of the game. If a team violates one of the rules of the game then the officials call a *penalty*. Penalties can be categorized as either a *dead ball penalty*, where the play is not allowed to continue and must be repeated, or a *live ball penalty*, where the play is allowed to continue to completion.[9] For our purposes, the distinction is important in that we can observe the type of play called on a live ball penalty, but not on a dead ball penalty. Additionally, we note that the game is divided into four quarters of 15 minutes. Play is stopped at the end of the second quarter and in the third quarter the game is restarted under different conditions than those at the end of the second quarter. The game ends at the end of the fourth quarter. Therefore, we refer to the end of the second and fourth quarters as the *end of play*.

1.2 Related Literature

We have seen that laboratory studies do not find strong evidence that subjects play mixed strategies according to minimax. However, these studies are vulnerable to the criticism that laboratory subjects face relatively small material incentives and do not possess the experience necessary to adequately mix. In response to this critique, there is a growing literature that examines mixed strategies in a sports setting. These studies have the advantage that the participants face large material incentives for success and the participants have a great deal of experience in these settings.[10]

[8] We acknowledge that, while the coaches are one source of the called play, it is also the case that many teams allow the offensive players to change the play after viewing the alignment of the defense. Our data set does not allow us to distinguish between these possibilities. Therefore, we simply regard the decision making unit as the team.

[9] For live ball penalties, the offended team generally has the option to either accept the penalty and replay the down, or to decline the penalty and accept the result of the play. Below, we refer to accepted live ball penalties as simply live ball penalties.

[10] Goldman and Rao (2013) find that professional basketball players are largely successful at solving the complex optimization problem regarding the decision to shoot or wait for a better shot before the time in which the team is required to shoot.

Walker and Wooders (2001) examine the direction of serves in professional tennis matches. The authors find that the probability of success for serves to the right and serves to the left are not different, which is consistent with the predictions of minimax. However, the authors note that the serves exhibit negative serial correlation, whereby the direction of a serve is not independent of the direction of the previous serve. Hsu, Huang, and Tang (2007) preform a similar analysis on a different tennis data set. In contrast to Walker and Wooders (2001), the authors do not find evidence of serial correlation of serves.

Other papers examine the direction of penalty kicks in soccer. The evidence largely supports the contention that the participants mix according to minimax (Chiappori, Levitt, and Groseclose, 2002; Palacios-Huerta, 2003; Coloma, 2007; Azar and Bar-Eli, 2011; Buzzacchi and Pedrini, 2014). In contrast, Bar-Eli et al. (2007) examine the behavior of soccer goalkeepers in a penalty kick, where the action choice is either to dive to the left, dive to the right, or stay in the middle of the goal. The authors find that the frequency with which goalkeepers stay in the middle is excessively small. The authors interpret this as an Action Bias, whereby the goalkeepers have a preference to be perceived as doing *something* to attempt to keep the goal from being scored, despite that this is suboptimal for the purposes of preventing the goal.[11]

To our knowledge, there are two previous studies that investigate serial correlation in the pass-rush mix in football, Kovash and Levitt (2009) and McGarrity and Linnen (2010). McGarrity and Linnen (2010) examine play calling while restricting attention to first downs with a distance of ten yards. Their data is taken from 11 NFL teams during the 2006 season. The authors perform a perform a test of runs for serial independence. Based on their analysis, the authors can only reject serial independence for one of the 11 teams. The authors argue that their data support the claim that play calling largely does not exhibit serial correlation. By contrast, the analysis of our more extensive data is consistent with the claim that there is serial correlation in play calling and that it leads to plays with lower efficacy.

[11] Another line of research investigates whether the ability to mix according to minimax in a familiar strategic setting in the field translates to the ability to mix properly in an unfamiliar strategic setting in the laboratory. We note that the conclusions in this literature are not uncontroversial (see Levitt, List, and Reiley, 2010; Palacios-Huerta and Volij, 2008; Van Essen and Wooders, 2013; Wooders, 2010).

As we do, Kovash and Levitt (2009) find negative serial correlation across plays. We differ from these authors in that we employ different measures of the efficacy of a play. Kovash and Levitt estimate the expected number of points, given any profile of down, distance, and field position. Their measure of efficacy entails calculating the difference in the expected points before and after every play. By contrast, our measures are more standard (yards gained and whether the play was successful according to a standard measure) and we explicitly control for the profile of down, distance, and field position in our econometric specification. We favor the measures which we use over the expected points measure because the "true value" of the latter will vary by team, by year, and even by the available personnel. There are additional differences as Kovash and Levitt compare the efficacy of a rush and a pass, and conclude that the play calling violates minimax. By contrast, our examination of minimax does not compare the differences in outcomes between a pass and a rush, but rather we study whether the efficacy of plays are affected by previous outcomes in a manner that is not consistent with minimax.

2 Data

2.1 Overview

Our data was obtained from http://armchairanalysis.com for a small fee. The data was taken from each regular season and playoff game from the 2000 season through the 2012 season. The data set contains $562,564$ plays from 3455 games. As is standard in the literature that examines football data,[12] we restrict attention to situations in which the game is neither affected by the end of play nor affected by a large score differential. Therefore, we exclude from our analysis plays that occurred in the last 2 minutes of the second quarter, plays that occurred in the fourth quarter,[13] and plays that occurred when the absolute value of the point difference[14] was 22 points or greater. After also excluding non-scrimmage plays (kickoffs,

[12] For instance, see Romer (2006) and Kovash and Levitt (2009).

[13] If the game is tied at the end of the fourth quarter, the teams go on to play an additional period referred to as overtime. We also exclude plays that occurred in overtime.

[14] We calculate this by subtracting the score of the defense from the score of the offense.

punts, field goal attempts, extra points, and two-point conversions), we have $257,782$ plays from scrimmage and $267,584$ offensive play decisions.

From a brief description of the play, we categorize each play as either a rush or a pass. Most of our categorizations should not be controversial and are identical to that provided by the data set. We do, however, categorize a "lateral pass" as a pass, whereas the official records categorize this as a rush. Further, we categorize any play in which an illegal forward pass penalty is called as a pass and not a rush. Finally, we categorize quarterback sacks as a pass, since the play is a failed pass play.

In addition to the down, distance, and field position of each play, our data includes the home team, the current score, the betting point spread, and whether there was a penalty on the play. Further, our data includes the conditions of the game: a characterization of the weather and wind conditions, and whether the game was played on grass or artificial turf.

We also include two different measures of the efficacy of a play. The first measure we use is the number of yards gained by the play. The second measure we use is whether the play is successful according to the standard measure: if on 1st down, 40% of the distance is gained, on 2nd down, 60% of the distance is gained, and on 3rd and 4th downs, 100% of the distance is gained.[15] Finally, we define a play as a *failure* if one yard or less is gained.

2.2 Different specifications of the previous play

As this paper explores serial correlation in NFL play calling, the classification of the *previous play* is crucial. First we note that if a play is the first of a possession then we do not assign it a previous play. However, if the play is not the first of a possession then we possibly assign it a previous play. Due to the nature of penalties in football, there are several ways to determine the previous play. In the remainder of this subsection, we carefully describe the four that we employ in the analysis. We employ four distinct definitions since we want our results to be robust to the precise definition of the previous play. The reader who is not interested in these details can skip the remainder of this subsection, keeping in mind that we employ four distinct definitions, each of which imply a different number of observations.

[15] This standard measure was included in the original data set.

One way to consider the previous play is based on whether there is a dead ball penalty on the play or on the proceeding plays. If a dead ball penalty is called on a play then the play is not assigned a previous play. If it is neither the case that a dead ball penalty is called on the play nor called on the proceeding play then we assign the type given to the proceeding play. If there was a dead ball penalty in the proceeding play but not the play proceeding that, then we use the play type assigned to the play before the penalty. If the two proceeding plays involved dead ball penalties then we assign the play type given to the play proceeding the first dead ball penalty. We continue in this manner for any number of consecutive dead ball penalties. We therefore include the information obtained from plays in which a live ball penalty was called and the teams consider the information learned prior to the dead ball penalties. We describe this specification as *Previous 1.* There are $209,963$ plays from scrimmage with an observation involving Previous 1.

Another way to consider the previous play is done as in Previous 1, except that plays following dead ball penalties are not assigned previous plays. In other words, here the teams do not consider information prior to a dead ball penalty. We describe this specification as *Previous 2.* There are $203,791$ plays from scrimmage with an observation involving Previous 2.

An additional way to consider the previous play is done as Previous 1, except that information observed in plays with a live ball penalty is not used. In other words, if it is neither the case that a penalty was called on the play nor called on the proceeding play then we assign the type given to the proceeding play. If there was a penalty in the proceeding play but not the play proceeding that, then we use the play type assigned to the play before the penalty. If the two proceeding plays involved penalties then we simply assign the play type given to the play proceeding the first penalty. We continue in this manner for any number of consecutive penalties. We describe this variable as *Previous 3.* There are $202,329$ plays from scrimmage with an observation involving Previous 3.

Finally, one could consider the previous play as in Previous 3, except that plays following penalties are not assigned previous plays. In other words, here the teams do not consider the information learned prior to a penalty. We describe this variable as *Previous 4*. There are 194,860 plays from scrimmage with an observation involving Previous 4.

We summarize the differences among these four techniques in Table 1, where a check indicates that it satisfies the criteria.

Table 1 Summary of the differences among previous play classifications

	Previous 1	Previous 2	Previous 3	Previous 4
1. Include live ball penalties	✓	✓		
2. Include plays following live ball penalties	✓	✓	✓	
Following a live ball penalty:				
2a. previous play is the most recent live ball penalty	✓	✓		
2b. previous play is the most recent non-penalized play			✓	
3. Include plays following dead ball penalties	✓		✓	

To illustrate the differences between the previous classifications, consider the following example. The first play of the possession is a pass. The second is a rush. The third is a live ball penalty on a pass play. The fourth play is a rush. The fifth play is a dead ball penalty. The sixth and seventh plays are passes. Table 2 illustrates the differences in the previous classifications in this example.

Table 2 An example sequence of plays and the corresponding previous classifications

Play	Play Type	Previous 1	Previous 2	Previous 3	Previous 4
1	Pass	–	–	–	–
2	Rush	*Pass*	*Pass*	*Pass*	*Pass*
3	Pass-Live ball penalty	*Rush*	*Rush*	–	–
4	Rush	*Pass*	*Pass*	*Rush*	–
5	No play-Dead ball penalty	–	–	–	–
6	Pass	*Rush*	–	*Rush*	–
7	Pass	*Pass*	*Pass*	*Pass*	*Pass*

Note that the difference between Previous 1 and 3 lies in whether information about the play call for a live ball penalty is considered. The difference between Previous 1 and 2 lies in whether information prior to a dead ball penalty is considered. The difference between

Previous 3 and 4 lies in whether information prior to any penalty is considered. Finally, note that in the fourth play of the possession, Previous 1 and 2 have a different assignment than Previous 3. This is because Previous 1 and 2 consider information that the previous play, which was given a live ball penalty, was a pass play. In contrast, Previous 3 disregards the information of the play call in the live ball penalty but considers information learned prior to the play call in the live ball penalty. Therefore, Previous 3 categorizes the previous play as a rush and not a pass.

3 Results

3.1 Summary statistics

In the analysis below, our independent variables include the down, the distance, the field position,[16] the point difference, and the difference between the point difference and the betting point spread. In order to account for the particular matchup between the teams, we include the fraction of plays that were passes by the offense within the particular game, the yards per pass earned within the game, the yards per rush earned within the game, and the fraction of plays within the game that were considered a success. We offer a summary of several key independent variables in Table 3.

Table 3 Summary statistics

	Mean	SD	Min	Max
Down	1.786	0.800	1	4
Distance	8.584	3.833	1	48
Fraction of pass plays in game	0.562	0.112	0.123	0.891
Fraction of successful plays in game	0.447	0.082	0.0952	0.773
Yards earned per pass in game	6.140	1.951	−0.500	19.818
Yards earned per rush in game	4.046	1.270	−2.375	13.562
Point difference	−0.790	7.786	−21	21
Play was a failure	0.380	0.485	0	1

We list the summary statistics for several key independent variables. The mean and standard deviation calculations are performed on the play-level rather

[16] We treat this as a categorical variable indicating whether the play originated 81 or more yards from their goal, between 51 and 80 yards, between 50 and 21 yards, between 20 and 6, or 5 yards or less.

than the game-level. The data includes 257,782 observations. We note that these calculations include plays which do not have a *previous* play.

We also describe the differences between pass plays and rush plays. We use two measures of these types of plays: yards gained by the play and whether the play was successful. We summarize this comparison in Table 4.

Table 4 Comparison between pass and rush

	Yards		Successful	
	Mean	SD	Mean	SD
Pass	6.212	10.158	0.443	0.497
Rush	4.264	6.357	0.458	0.498
z-statistic	-7.57		8.09	
p-value	< 0.001		< 0.001	

We provide the mean and standard deviation of both the yards gained and whether the play was a success, by play type. We also report the results of Mann-Whitney tests of the difference between rush and pass plays. The data includes 257,782 observations, involving 139,302 pass plays and 118,480 rush plays. Note that these calculations include plays which do not have a *previous* play.

We note that pass plays, on average, gain more yards than rush plays, though rush plays more often satisfy our definition of a successful play.[17] These two differences are significant according to Mann-Whitney tests. We also note that, while pass plays have a significantly larger mean of yards gained, they also have a larger standard deviation of yards gained as measured by an F-test of the equality of variances ($F(139301, 118479) = 2.55$, $p < 0.001$).

3.2 Serial correlation

We now move to the first of our primary research questions, whether the pass-rush mix exhibits serial correlation. Our dependent variable is a dummy variable that takes a value of 1 if the play is a pass, and a 0 otherwise. Due to the binary nature of our dependent variable, we run logistic regressions. Our independent variables include the down, the distance, the field

[17] For more on the optimality of the fraction of rush plays and pass plays, see Alamar (2006, 2010), Reed, Critchfield, and Martens (2006), Rockerbie (2008), Kovash and Levitt (2009), and Stilling and Critchfield (2010).

position, the point difference, the difference between the point difference and the betting point spread, the fraction of plays that were passes by the offense within the particular game, the yards per pass earned within the game, the yards per rush earned within the game, and the fraction of plays within the game that were considered a success. Additionally, we account for various observables, such as whether the game was played in excessively cold conditions,[18] excessively windy conditions,[19] wet conditions,[20] whether the game was played on grass,[21] and whether the offense was also the home team. We include team-season fixed-effects in our regressions. We run four regressions, one for each of our four techniques for determining the previous play. We summarize this analysis in Table 5.

Table 5 Logistic regressions of serial correlation: Play is a pass

	(1)	(2)	(3)	(4)
Previous pass	-0.312^{***}	-0.318^{***}	-0.317^{***}	-0.330^{***}
	(0.00558)	(0.00567)	(0.00567)	(0.00581)
Previous failure	0.00495	-0.00848	-0.0217^{***}	-0.0564^{***}
	(0.00735)	(0.00761)	(0.00738)	(0.00798)
Previous pass * Previous failure	-0.0998^{***}	-0.106^{***}	-0.111^{***}	-0.120^{***}
	(0.0055)	(0.0056)	(0.00559)	(0.00575)
-2 Log L	251135.17	243935.06	241845.61	232354.56
LR χ^2	36918.69^{***}	35950.42^{***}	35952.54^{***}	35187.45^{***}
Observations	209,963	203,791	202,329	194,860

We do not list the estimates of the other independent variables, the estimate of the intercept, or the estimates of the team-season fixed-effects. Note that * indicates significance at $p < 0.1$, ** indicates significance at $p < 0.01$, and *** indicates significance at $p < 0.001$.

First, we note that in each of the four specifications, the previous pass variable is negative and significant. This suggests that a pass is less likely to be called following a pass play. We also note that the Previous pass-Previous failure interaction estimate is negative and

[18] We have two categories: if the temperature is less than 20 Fahrenheit (-6.67 Celsius) or if it is greater than 20 degrees Fahrenheit but less than 30 Fahrenheit (-1.1 Celsius).

[19] We have two categories: if the wind speed is higher than 30 miles per hour (mph) or if it is less than 30 mph but higher than 20 mph.

[20] We note whether the description of the game included a mention of snow, rain, or flurries.

[21] See Bailey and McGarrity (2012) for an example of an analysis which also considers the playing surface. Unlike these authors, we do not find a significant effect.

significant in each of the four specifications. This suggests that a pass is particularly less likely to be called following a failed pass play.

We note that this negative serial correlation is not consistent with the predictions of minimax. For example, after controlling for all other factors, we find that a pass is less likely if the previous play was a pass, especially if the previous play failed. However, it remains to be seen whether this negative serial correlation affects the efficacy of the plays.

3.3 Serial correlation and yards gained

Above, we find negative serial correlation in the pass-rush mix. However, it is important to discern whether this pattern has an appreciable impact on the efficacy of those plays. As we do not have data on the strategy of the defensive team, we examine whether the observed negative serial correlation is a best response to the unobserved strategy of the defense.

We begin our analysis with yards gained as the dependent variable. We include dummy variables indicating whether the play is a pass, whether the previous play was a failure, and whether the play type is the same type (rush or pass) as the previous play. As described above, we use four methods to specify the previous play. As in the analysis summarized by Table 5, we include team-seasons fixed-effects. In addition to these variables, we also include the control variables that were used in the analysis summarized in Table 5. We summarize this analysis in Table $A1$, in the appendix. Given the analysis summarized in Table $A1$, we estimate various differences which most interest us and conduct Wald tests on these estimates. In particular, we examine the extent to which differences in previous actions and previous outcomes affect the yards gained by a play of a particular type: rush or pass. In other words, we do not compare the differences in the yards gained on a pass play and a rush play, rather we restrict attention to a particular play type and investigate whether past actions and past outcomes affect the yards gained. This analysis and the regression statistics from Table $A1$ are summarized in Table 6.

Table 6 Estimates of the difference in yards gained between a

	(1)	(2)	(3)	(4)
Rush following a rush and	0.138^{*}	0.134^{*}	0.129^{*}	0.148^{*}
a rush following a pass	(0.0674)	(0.0683)	(0.0679)	(0.0698)
Pass following a pass and	0.211^{***}	0.234^{***}	0.206^{***}	0.236^{***}
a pass following a rush	(0.0511)	(0.0520)	(0.0522)	(0.0532)
Rush following a failed rush and	0.238^{*}	0.227^{*}	0.211^{*}	0.249^{*}
a rush following a failed pass	(0.118)	(0.120)	(0.119)	(0.123)
Rush following a non-failed rush and	0.0368	0.0413	0.0469	0.0472
a rush following a non-failed pass	(0.0646)	(0.0651)	(0.0654)	(0.0666)
Pass following a failed pass and	0.0834	0.107	0.0642	0.0771
a pass following a failed rush	(0.0776)	(0.0790)	(0.0794)	(0.0808)
Pass following a non-failed pass and	0.338^{***}	0.362^{***}	0.348^{***}	0.394^{***}
a pass following a non-failed rush	(0.0660)	(0.0670)	(0.0670)	(0.0685)
R^2	0.05	0.05	0.05	0.05
F-value	24.89^{***}	24.42^{***}	24.23^{***}	23.28^{***}
Observations	$209,963$	$203,791$	$202,329$	$194,860$

These comparisons are based on the analysis summarized in Table $A1$. Based on the Wald test of the estimates, * indicates significance at $p < 0.1$, ** indicates significance at $p < 0.01$, and *** indicates significance at $p < 0.001$.

First we note that there are no negative estimates in Table 6. In other words, given any of the situations which we consider, there is no evidence that running the same play type as the previous play presents a disadvantage in terms of yards gained. The first two rows only consider the play type and not whether the previous play was a failure. We find that a rush play following a rush play gains $0.13 - 0.15$ yards more than a rush play following a pass play. We also find that a pass play following a pass play gains $0.21 - 0.24$ yards more than a pass play following a rush play. We note that the estimates can be more pronounced when we also consider whether the previous play was a failure. We find that a rush play gains $0.21 - 0.25$ yards more following a failed rush play than following a failed pass play. We also find that a pass play gains $0.34 - 0.39$ yards more following a non-failed pass play than following a non-failed rush play. To summarize, we find evidence that the efficacy of a play, as measured by yards gained, increases if it follows a play of the same type. This is not consistent with minimax.

3.4 Serial correlation and success

Whereas we previously examined whether the serial correlation affected the yards gained by a play, it is possible that this measure does not completely capture the efficacy of a play. Therefore, we perform an analysis similar to that summarized in Table $A1$, however we use a dependent variable which assumes a value of 1 if the play is successful, and a 0 otherwise. Since the dependent variable is binary, we employ logistic regressions. We include the identical set of independent variables as the analysis summarized in Table $A1$. This analysis is summarized in Table $A2$, in the appendix. Given the analysis summarized in Table $A2$, we estimate the differences which most interest us and conduct Wald tests on these estimates. These comparisons are done in the same manner as Table 6. We summarize these comparisons and present the the regression statistics from Table $A2$ in Table 7.

Table 7 Difference in success estimates between a

	(1)	(2)	(3)	(4)
Rush following a rush and	0.123^{***}	0.127^{***}	0.112^{***}	0.114^{***}
a rush following a pass	(0.0176)	(0.0178)	(0.0177)	(0.0182)
Pass following a pass and	0.0274^{*}	0.0320^{*}	0.0279^{*}	0.0374^{**}
a pass following a rush	(0.0131)	(0.0133)	(0.0133)	(0.0136)
Rush following a failed rush and	0.251^{***}	0.258^{***}	0.240^{***}	0.246^{***}
a rush following a failed pass	(0.0313)	(0.0318)	(0.0314)	(0.0324)
Rush following a non-failed rush and	-0.0055	-0.0041	-0.0169	-0.0172
a rush following a non-failed pass	(0.0160)	(0.0162)	(0.0162)	(0.0165)
Pass following a failed pass and	0.0110	0.0139	0.0094	0.0203
a pass following a failed rush	(0.0202)	(0.0206)	(0.0207)	(0.0210)
Pass following a non-failed pass and	0.0437^{**}	0.0500^{**}	0.0465^{**}	0.0545^{**}
a pass following a non-failed rush	(0.0165)	(0.0167)	(0.0167)	(0.0171)
-2 Log L	269807.43	262498.12	260592.96	251358.09
Observations	209,963	203,791	202,329	194,860

These comparisons are based on the analysis summarized in Table $A2$. Based on the Wald test of the estimates, * indicates significance at $p < 0.1$, ** indicates significance at $p < 0.01$, and *** indicates significance at $p < 0.001$.

We note that there are no significant and negative estimates in Table 7, again suggesting that plays are more successful when they follow plays of the same type. Additionally, as we found earlier, our results become more pronounced when we condition on the possibility that the previous play was a failure. We find that a rush play is significantly more likely to be successful when following a failed rush play than when following a failed pass play. Additionally, we find that a pass play following a non-failed pass play is significantly more likely to be successful than a pass play following a non-failed rush play. Similar to Table 6, the analysis summarized in Table 7 provides evidence that the efficacy of plays are affected by previous actions and previous outcomes.

3.5 Analysis of second down

In the analysis above, we examined plays that occurred on each of the four downs. However, it is possible that behavior is sufficiently different across downs, so we analyze downs individually. We begin with second down, as second down always follows a particular down (first down) and there is a variation in the failure of the previous play. We conduct an analysis identical to that summarized in Table 5, restricting our attention to second down. We summarize this analysis in Table 8.

Table 8 Logistic regressions of serial correlation: Play is a pass on second down

	(1)	(2)	(3)	(4)
Previous pass	-0.485^{***}	-0.492^{***}	-0.493^{***}	-0.503^{***}
	(0.00819)	(0.00830)	(0.00831)	(0.00843)
Previous failure	-0.0550^{***}	-0.0588^{***}	-0.0556^{***}	-0.0787^{***}
	(0.0107)	(0.0110)	(0.0108)	(0.0114)
Previous pass * Previous failure	-0.157^{***}	-0.161^{***}	-0.169^{***}	-0.176^{***}
	(0.00806)	(0.00817)	(0.00817)	(0.00830)
-2 Log L	107882.13	105333.39	105310.64	102911.80
LR χ^2	12083.64^{***}	11802.50^{***}	11818.37^{***}	11642.38^{***}
Observations	86,645	84,574	84,569	82,683

We do not list the estimates of the other independent variables, the estimate of the intercept, or the estimates of the team-season fixed-effects. Note that * indicates significance at $p < 0.1$, ** indicates significance at $p < 0.01$, and *** indicates significance at $p < 0.001$.

First, we note that the Previous pass coefficients are negative and significant at 0.001, implying that a pass is significantly less likely to be called following a pass. Therefore, similar to the analysis summarized in Table 5, here we find evidence of negative serial correlation.

Also similar to Table 5, we find that the Previous pass-Previous failure interaction is negative and significant. This suggests that a pass on second down is even less likely after a failed pass on first down. However, unlike what was found in Table 5, here we find that the Previous failure coefficients are negative and significant at 0.001.

Next we investigate whether the negative serial correlation on second down effects the efficacy of plays. We conduct an analysis, similar to that summarized in Table $A2$, including only second down. This analysis is summarized in Table $A3$, in the appendix. Also similar to the analysis summarized in Table 6, here we estimate various differences found in the analysis. This analysis and the regression statistics from Table $A3$ are summarized in Table 9.

Table 9 Estimates of the difference in yards gained on second down between a

	(1)	(2)	(3)	(4)
Rush following a rush and	0.239^{**}	0.243^{**}	0.237^{**}	0.246^{**}
a rush following a pass	(0.0857)	(0.0865)	(0.0870)	(0.08742)
Pass following a pass and	0.452^{***}	0.459^{***}	0.443^{***}	0.441^{***}
a pass following a rush	(0.0851)	(0.0865)	(0.0860)	(0.0872)
Rush following a failed rush and	0.389^{**}	0.394^{**}	0.376^{**}	0.399^{**}
a rush following a failed pass	(0.1328)	(0.134)	(0.136)	(0.137)
Rush following a non-failed rush and	0.0886	0.0913	0.0985	0.0924
a rush following a non-failed pass	(0.107)	(0.108)	(0.108)	(0.108)
Pass following a failed pass and	0.315^{**}	0.347^{**}	0.288^{**}	0.312^{**}
a pass following a failed rush	(0.108)	(0.110)	(0.111)	(0.113)
Pass following a non-failed pass and	0.589^{***}	0.572^{***}	0.598^{***}	0.571^{***}
a pass following a non-failed rush	(0.131)	(0.133)	(0.131)	(0.133)
R^2	0.05	0.05	0.05	0.05
F-value	10.38^{***}	10.19^{***}	10.18^{***}	10.07^{***}
Observations	$86,645$	$84,574$	$84,569$	$82,683$

These comparisons are based on the analysis summarized in Table $A3$. Based on the Wald test of the estimates, * indicates significance at $p < 0.1$, ** indicates significance at $p < 0.01$, and *** indicates significance at $p < 0.001$.

Similar to that found in Table 6, we do not find a significant and negative estimate of the difference in yards gained on second down. Indeed, comparing Table 6 to Table 9 we see that the latter either has larger coefficient estimates or lower p-values of the estimates. Whereas Table 6 only has 2 of the 6 conditions that are significant at 0.01 in all four specifications, Table 9 has 5 of the 6 situations that are significant at 0.01 in all four specifications. We also find that a rush play following a rush play gains $0.24 - 0.25$ yards more than a rush play following a pass play. Additionally, a pass play following a pass play gains $0.44 - 0.46$ more yards than a pass play following a rush play. As in the previous analyses, estimates can become more pronounced when we also consider whether the previous play was a failure. A rush play gains $0.38 - 0.40$ yards more following a failed rush play than following a failed pass play. Further, a pass play gains $0.57 - 0.60$ yards more following a non-failed pass play than following a non-failed rush play. We also conduct the analogous analysis consisting of the success measure, rather than the yards gained measure, on second down plays.[22] Similar to Table 7, we do not find a negative and significant estimate, except in a single condition.[23] In summary, we find evidence of serial correlation of plays called on second down and that second down plays are more successful following plays of the same type. Again, this is not consistent with minimax.

3.6 Analysis of third down

We now examine whether there is serial correlation on third down plays. Like second down, the down before third down is known (second down) and there is variation in the failure of the previous play. This analysis is conducted in the same manner as that summarized in Table 8, with the exception that we restrict attention to third down plays. We summarize this analysis in Table 10.

[22] This is available from the corresponding author upon request.

[23] We find that there is a negative difference in the estimate of the probability of success of a rush following a non-failed rush and the probability of success of a rush following a non-failed pass, which is significant at 0.1 in all four specifications.

Table 10 Logistic regressions of serial correlation: Play is a pass on third down

	(1)	(2)	(3)	(4)
Previous pass	-0.0220^{*}	-0.0293^{*}	-0.0264^{*}	-0.0303^{*}
	(0.0117)	(0.0120)	(0.0120)	(0.0122)
Previous failure	0.0797^{***}	0.0674^{***}	0.0711^{***}	0.0605^{***}
	(0.0125)	(0.0129)	(0.0129)	(0.0130)
Previous pass * Previous failure	0.00765	-0.00191	0.000095	-0.00455
	(0.0117)	(0.0120)	(0.0120)	(0.0121)
-2 Log L	50204.85	48088.12	48060.55	47060.99
LR χ^2	6648.84^{***}	6905.33^{***}	6929.35^{***}	7033.16^{***}
Observations	$54,922$	$52,815$	$52,813$	$51,862$

We do not list the estimates of the other independent variables, the estimate of the intercept, or the estimates of the team-season fixed-effects. Note that * indicates significance at $p < 0.1$, ** indicates significance at $p < 0.01$, and *** indicates significance at $p < 0.001$.

Here we find some evidence of negative serial correlation on third down plays, although these estimates are significant only at 0.1. However, we find some differences from the previous analyses of serial correlation found in Tables 5 and 8. First, the Previous pass-Previous failure interaction is not significant in any of our specifications. Further, the Previous failure coefficient is positive and significant in all for specifications. In other words, when the previous play on second down was a failure, a pass is significantly more likely to be called on third down.

While we find some evidence of negative serial correlation among play calling on third down, we now investigate whether this serial correlation affects the efficacy of third down plays. We conduct an analysis similar to that in the previous subsection, however we restrict attention to third down plays. The analysis is summarized in Table $A4$, in the appendix. We also estimate the relevant differences, as in the analysis summarized in Tables 6 and 9. We summarize this analysis, in addition to the regression statistics from Table $A4$, in Table 11.

Table 11 Estimates of the difference in yards gained on third down between a

	(1)	(2)	(3)	(4)
Rush following a rush and	−0.101	−0.131	−0.133	−0.135
a rush following a pass	(0.180)	(0.182)	(0.183)	(0.185)
Pass following a pass and	0.0331	0.0520	0.0260	0.0464
a pass following a rush	(0.0940)	(0.0957)	(0.0953)	(0.0964)
Rush following a failed rush and	−0.324	−0.390	−0.406	−0.397
a rush following a failed pass	(0.279)	(0.285)	(0.287)	(0.290)
Rush following a non-failed rush and	0.122	0.128	0.140	0.128
a rush following a non-failed pass	(0.227)	(0.228)	(0.227)	(0.228)
Pass following a failed pass and	−0.141	−0.115	−0.149	−0.128
a pass following a failed rush	(0.124)	(0.127)	(0.127)	(0.129)
Pass following a non-failed pass and	0.207	0.219	0.201	0.221
a pass following a non-failed rush	(0.140)	(0.142)	(0.141)	(0.143)
R^2	0.05	0.06	0.06	0.06
F-value	7.17***	7.02***	7.02***	6.90***
Observations	54,922	52,815	52,813	51,862

These comparisons are based on the analysis summarized in Table $A4$. Based on the Wald test of the estimates, * indicates significance at $p < 0.1$, ** indicates significance at $p < 0.01$, and *** indicates significance at $p < 0.001$.

We note that unlike Tables 6 and 9, Table 11 does not contain a single significant relationship. In other words, the weak evidence of negative serial correlation found in Table 10 does not appear to affect the efficacy of third down plays. We also note that the analogous analysis consisting of the success of the play, rather than the yards gained, is qualitatively similar to that of Table 11.[24] Given the lack of evidence of an effect of the serial correlation on third down plays, the results unrestricted by down, found in Tables 6 and 7, are all the more surprising.

4 Conclusion

We analyze play calling in 3455 National Football League games from the 2000 season through the 2012 season. We categorize every relevant play as either a pass or a rush. We find that the play calling exhibits negative serial correlation. In other words, we find evidence that

[24] The one exception is that there is a positive difference in the estimate of the probability of success of a rush following a failed rush and the probability of success of a rush following a failed pass, which is significant at 0.1 in all four specifications. This is available from the corresponding author upon request.

the type of play called is affected by the previous type of play and that the play types switch more than that implied by an independent stochastic process.

We also find that the efficacy of plays are affected by previous actions and previous outcomes in a manner that is not consistent with minimax. In particular, we find that a rush play following a rush play earns more yards and is more likely to be successful than a rush play following a pass play. Similarly, we find that a pass play following a pass play earns more yards and is more likely to be successful than a pass play following a rush play. Given that the yards and success measures exhibit very different properties, it should be all the more surprising that their qualitative implications regarding the effects of serial correlation are similar. Further, we find that these differences can become more pronounced when we also consider whether the previous play was a failure.

We also conduct an analysis of serial correlation separately by down. We find evidence that plays on second down exhibit negative serial correlation and that the effects of the serial correlation are pronounced. However, when we restrict attention to third down, we only find weak evidence of negative serial correlation and we do not find evidence that this serial correlation affects the outcomes of third down plays. Therefore, it seems that our results are driven by second down behavior, and not third down behavior.

What are we to make of our results? How could it be the case that experienced decision makers under large material incentives for success, with the ability to make detailed plans prior to the decision, and who could consult others in the decision exhibit behavior that is not consistent with minimax?

A possible explanation for our findings of serial correlation is that the play calling switches too often because of the fatigue of the players involved in the play. This is a possible explanation of the negative serial correlation of plays but it does not explain the results regarding the efficacy of plays. In particular, we find that there is a positive benefit when two consecutive plays are of the same type rather than when the two consecutive plays are different types. Therefore, the material effects of fatigue cannot explain our results. It is possible that differences in the perception of fatigue by the offense and defense could explain our results. However, it is difficult for us to see how these differences in the perception of fatigue could be sufficiently and systematically different to explain our results.

Another possible explanation follows from the research that indicates that people have difficulty producing independent, random sequences.[25] Whereas this could explain the negative serial correlation of play calling, it cannot explain the reduced efficacy of plays associated with the serial correlation. The latter is consistent with the claim that the defense expects the play calling of the offense to excessively switch play type.

It is possible that there are excessive computational difficulties in accurately mixing. The teams must not simply decide to rush or to pass but rather which of the several hundred pass or rush plays to execute. Perhaps the effects of these computational difficulties could explain our results.[26] While this can explain the negative serial correlation of play calling, it is not clear why the lower efficacy associated with the negative serial correlation would be so persistent.

Finally, it is possible that the teams feel pressure not to repeat the play type on offense, in order to avoid criticism for being too "predictable" by people who have difficulty detecting whether outcomes of a sequence are statistically independent. Further, perhaps this concern is sufficiently important so that teams accept the negative consequences that arise from the risk that the defense can detect a pattern in their mixing.[27] This explanation is reminiscent of the Action bias found by Bar-Eli et al. (2007). The explanation that teams do not want to be viewed as predictable and accept the reduction in the efficacy of their plays which result from the negative serial correlation seems to be the explanation most consistent with our data.

---

[25] See Wagenaar (1972), Bar-Hillel, and Wagenaar (1991), Rabin (2002), and Oskarsson, Van Boven, McClelland, and Hastie (2009).

[26] For instance, see Halpern and Pass (2014).

[27] See Shachat and Swarthout (2004) and Spiliopoulos (2012).

Appendix

Table A1 Regressions of yards gained

	(1)	(2)	(3)	(4)
Pass	1.715***	1.729***	1.738***	1.724***
	(0.0639)	(0.0646)	(0.0648)	(0.0662)
Previous failure	0.470***	0.449***	0.439***	0.419***
	(0.0964)	(0.0982)	(0.0963)	(0.101)
Same play type as previous	0.0368	0.0413	0.0469	0.0472
	(0.0646)	(0.0651)	(0.0654)	(0.0666)
Pass * Previous failure	−0.515***	−0.531***	−0.506***	−0.501***
	(0.112)	(0.114)	(0.113)	(0.116)
Previous failure * Same	0.202	0.186	0.164	0.201
	(0.135)	(0.137)	(0.136)	(0.140)
Pass * Same	0.301**	0.321***	0.301***	0.347***
	(0.0951)	(0.0962)	(0.0961)	(0.0982)
Pass * Previous failure * Same	−0.456**	−0.441*	−0.448**	−0.518**
	(0.171)	(0.173)	(0.173)	(0.177)
R^2	0.05	0.05	0.05	0.05
F-value	24.89***	24.42***	24.23***	23.28***
Observations	209,963	203,791	202,329	194,860

We do not list the estimates of the other independent variables, the estimate of the intercept, or the estimates of the team-season fixed-effects. Note that * indicates significance at $p < 0.1$, ** indicates significance at $p < 0.01$, and *** indicates significance at $p < 0.001$.

Table A2 Logistic regressions of a successful play

	(1)	(2)	(3)	(4)
Pass	0.0149	0.0097	0.0043	-0.0032
	(0.0159)	(0.0161)	(0.0162)	(0.0165)
Previous Failure	-0.291^{***}	-0.294^{***}	-0.281^{***}	-0.267^{***}
	(0.0254)	(0.0260)	(0.0254)	(0.0267)
Same play type as Previous	-0.0055	-0.0041	-0.0169	-0.0172
	(0.0160)	(0.0162)	(0.0162)	(0.0165)
Pass * Previous Failure	0.373^{***}	0.376^{***}	0.378^{***}	0.389^{***}
	(0.0293)	(0.0297)	(0.0295)	(0.0303)
Previous Failure * Same	0.256^{***}	0.262^{***}	0.257^{***}	0.263^{***}
	(0.0352)	(0.0357)	(0.0354)	(0.0364)
Pass * Same	0.0492^{*}	0.0541^{*}	0.0634^{**}	0.0717^{**}
	(0.0237)	(0.0240)	(0.0240)	(0.0245)
Pass * Previous Failure * Same	-0.289^{***}	-0.299^{***}	-0.294^{***}	-0.297^{***}
	(0.0442)	(0.0449)	(0.0446)	(0.0458)
-2 Log L	269807.43	262498.12	260592.96	251358.09
Observations	209,963	203,791	202,329	194,860

We do not list the estimates of the other independent variables, the estimate of the intercept, or the estimates of the team-season fixed-effects.. Note that * indicates significance at $p < 0.1$, ** indicates significance at $p < 0.01$, and *** indicates significance at $p < 0.001$.

Table A3 Regressions of yards gained on second down

	(1)	(2)	(3)	(4)
Pass	1.671***	1.686***	1.680***	1.686***
	(0.108)	(0.109)	(0.108)	(0.109)
Previous failure	0.209	0.205	0.207	0.200
	(0.131)	(0.134)	(0.132)	(0.136)
Same play type as previous	0.0886	0.0913	0.0985	0.0924
	(0.107)	(0.108)	(0.108)	(0.108)
Pass * Previous failure	−0.200	−0.225	−0.204	−0.210
	(0.1500)	(0.152)	(0.152)	(0.153)
Previous failure * Same	0.301*	0.302*	0.278	0.306*
	(0.170)	(0.172)	(0.173)	(0.174)
Pass * Same	0.501**	0.480**	0.499**	0.478**
	(0.171)	(0.173)	(0.171)	(0.173)
Pass * Previous failure * Same	−0.575*	−0.527*	−0.587*	−0.565*
	(0.242)	(0.245)	(0.245)	(0.248)
R^2	0.05	0.05	0.05	0.05
F-value	10.38***	10.19***	10.18***	10.07***
Observations	86, 645	84, 574	84, 569	82, 683

We do not list the estimates of the other independent variables, the estimate of the intercept, or the estimates of the team-season fixed-effects. Note that * indicates significance at $p < 0.1$, ** indicates significance at $p < 0.01$, and *** indicates significance at $p < 0.001$.

Table A4 Regressions of yards gained on third down

	(1)	(2)	(3)	(4)
Pass	1.050***	1.097***	1.107***	1.092***
	(0.201)	(0.203)	(0.202)	(0.203)
Previous failure	1.348***	1.359***	1.382***	1.376***
	(0.269)	(0.273)	(0.274)	(0.277)
Same play type as previous	0.122	0.128	0.140	0.127
	(0.227)	(0.228)	(0.227)	(0.228)
Pass * Previous failure	−1.208***	−1.286***	−1.297***	−1.287***
	(0.295)	(0.299)	(0.301)	(0.304)
Previous failure * Same	−0.446	−0.517	−0.546	−0.525
	(0.359)	(0.364)	(0.365)	(0.369)
Pass * Same	0.0852	0.0912	0.0609	0.0938
	(0.267)	(0.269)	(0.268)	(0.270)
Pass * Previous failure * Same	0.0975	0.184	0.196	0.175
	(0.407)	(0.414)	(0.414)	(0.418)
R^2	0.05	0.06	0.06	0.06
F-value	7.17***	7.02***	7.02***	6.90***
Observations	54,922	52,815	52,813	51,862

We do not list the estimates of the other independent variables, the estimate of the intercept, or the estimates of the team-season fixed-effects. Note that * indicates significance at $p < 0.1$, ** indicates significance at $p < 0.01$, and *** indicates significance at $p < 0.001$.

References

Alamar, Benjamin C. (2006): "The passing premium puzzle," *Journal of Quantitative Analysis in Sports*, 2(4), Article 5, 1-8.

Alamar, Benjamin C. (2010): "Measuring risk in NFL playcalling," *Journal of Quantitative Analysis in Sports*, 6(2), Article 11, 1–7.

Azar, Ofer H. and Bar-Eli, Michael (2011): "Do soccer players play the mixed-strategy Nash equilibrium?" *Applied Economics*, 43(25), 3591–3601.

Bailey, Brett James and McGarrity, Joseph P. (2012): "The Effect of Pressure on Mixed-Strategy Play in Tennis: The Effect of Court Surface on Service Decisions," *International Journal of Business and Social Science*, 3(20), 11–18.

Bar-Eli, Michael, Azar, Ofer H., Ritov, Ilana, Keidar-Levin, Yael, and Schein, Galit (2007): "Action bias among elite soccer goalkeepers: The case of penalty kicks," *Journal of Economic Psychology*, 28(5), 606–621.

Bar-Hillel, Maya, and Wagenaar, Willem A. (1991): "The perception of randomness," *Advances in Applied Mathematics*, 12(4), 428–454.

Batzilis, Dimitris, Jaffe, Sonia, Levitt, Steven, List, John A., and Picel, Jeffrey (2013): "How Facebook Can Deepen our Understanding of Behavior in Strategic Settings: Evidence from a Million Rock-Paper-Scissors Games," working paper, Harvard University.

Binmore, Ken, Swierzbinski, Joe, and Proulx, Chris (2001): "Does minimax work? An experimental study," *Economic Journal*, 111(473), 445–464.

Brown, James N. and Rosenthal, Robert W. (1990): "Testing the minimax hypothesis: a re-examination of O'Neill's game experiment," *Econometrica*, 58(5), 1065–1081.

Budescu, David V. and Rapoport, Amnon (1994): "Subjective randomization in one-and two-person games," *Journal of Behavioral Decision Making*, 7(4), 261–278.

Buzzacchi, Luigi and Pedrini, Stefano (2014): "Does player specialization predict player actions? Evidence from penalty kicks at FIFA World Cup and UEFA Euro Cup," *Applied Economics*, 46(10), 1067–1080.

Chiappori, P-A., Levitt, Steven, and Groseclose, Timothy (2002): "Testing mixed-strategy equilibria when players are heterogeneous: the case of penalty kicks in soccer," *American Economic Review*, 92(4), 1138–1151.

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

Geng, Sen, Peng, Yujia, Shachat, Jason, and Zhong, Huizhen (2014): "Adolescents, Cognitive Ability, and Minimax Play," working paper, Xiamen University

Goldman, Matt and Rao, Justin M. (2013): "Tick Tock Shot Clock: Optimal Stopping in NBA Basketball," working paper, UCSD.

Halpern, Joseph Y. and Pass, Rafael (2014): "Algorithmic rationality: Game theory with costly computation," *Journal of Economic Theory*, forthcoming.

Hsu, Shih-Hsun, Huang, Chen-Ying, and Tang, Cheng-Tao (2007): "Minimax play at wimbledon: comment," *American Economic Review*, 97(1), 517–523.

Kovash, Kenneth and Levitt, Steven D. (2009): "Professionals do not play minimax: evidence from major League Baseball and the National Football League," working paper, National Bureau of Economic Research.

Levitt, Steven D., List, John A., and Reiley, David H. (2010): "What happens in the field stays in the field: Exploring whether professionals play minimax in laboratory experiments," *Econometrica*, 78(4), 1413–1434.

McGarrity, Joseph P. and Linnen, Brian (2010): "Pass or Run: An Empirical Test of the Matching Pennies Game Using Data from the National Football League," *Southern Economic Journal*, 3, 791–810.

Mookherjee, Dilip and Sopher, Barry (1997): "Learning and decision costs in experimental constant sum games," *Games and Economic Behavior*, 19(1), 97–132.

Mookherjee, Dilip and Sopher, Barry (1994): "Learning Behavior in an Experimental Matching Pennies Game," *Games and Economic Behavior*, 7(1), 62–91.

Ochs, Jack (1995): "Games with unique, mixed strategy equilibria: An experimental study," *Games and Economic Behavior*, 10(1), 202–217.

Okano, Yoshitaka (2013): "Minimax play by teams," *Games and Economic Behavior*, 77(1), 168–180.

O'Neill, Barry (1987): "Nonmetric test of the minimax theory of two-person zerosum games," *Proceedings of the National Academy of Sciences*, 84(7), 2106–2109.

O'Neill, Barry (1991): "Comments on Brown and Rosenthal's reexamination," *Econometrica*, 59(2), 503–507.

Oskarsson, An T., Van Boven, Leaf, McClelland, Gary H., and Hastie, Reid (2009): "What's next? Judging sequences of binary events," *Psychological Bulletin*, 135(2), 262–285.

Palacios-Huerta, Ignacio (2003): "Professionals play minimax," *Review of Economic Studies*, 70(2), 395–415.

Palacios-Huerta, Ignacio and Volij, Oscar (2008): "Experientia docet: Professionals play minimax in laboratory experiments," *Econometrica*, 76(1), 71–115.

Rabin, Matthew (2002): "Inference by Believers in the Law of Small Numbers," *Quarterly Journal of Economics*, 117(3), 775–816.

Rapoport, Amnon and Amaldoss, Wilfred (2000): "Mixed strategies and iterative elimination of strongly dominated strategies: an experimental investigation of states of knowledge," *Journal of Economic Behavior and Organization*, 42(4), 483–521.

Rapoport, Amnon and Amaldoss, Wilfred (2004): "Mixed-strategy play in single-stage first-price all-pay auctions with symmetric players," *Journal of Economic Behavior and Organization*, 54(4), 585–607.

Rapoport, Amnon and Boebel, Richard B. (1992): "Mixed strategies in strictly competitive games: a further test of the minimax hypothesis," *Games and Economic Behavior*, 4(2), 261–283.

Rapoport, Amnon and Budescu, David V. (1992): "Generation of random series in two-person strictly competitive games," *Journal of Experimental Psychology: General*, 121(3), 352–363.

Reed, Derek D., Critchfield, Thomas S., and Martens, Brian K. (2006): "The generalized matching law in elite sport competition: Football play calling as operant choice," *Journal of Applied Behavior Analysis*, 39(3), 281–297.

Rockerbie, Duane W. (2008): "The passing premium puzzle revisited," *Journal of Quantitative Analysis in Sports*, 4(2), Article 9, 1–11

Romer, David (2006): "Do firms maximize? Evidence from professional football," *Journal of Political Economy*, 114(2), 340–365.

Rosenthal, Robert W., Shachat, Jason and Walker, Mark (2003): "Hide and seek in Arizona," *International Journal of Game Theory*, 32(2), 273–293.

Shachat, Jason M. (2002): "Mixed strategy play and the minimax hypothesis," *Journal of Economic Theory*, 104(1), 189–226.

Shachat, Jason and Swarthout, J. Todd (2004): "Do we detect and exploit mixed strategy play by opponents?" *Mathematical Methods of Operations Research*, 59(3), 359–373.

Spiliopoulos, Leonidas (2012): "Pattern recognition and subjective belief learning in a repeated constant-sum game," *Games and Economic Behavior*, 75(2), 921–935.

Stilling, Stephanie T. and Critchfield, Thomas S. (2010): "The Matching Relation and Situation-Specific Bias Modulation in Professional Football Play Selection," *Journal of the Experimental Analysis of Behavior*, 93(3), 435–454.

Van Essen, Matt, and Wooders, John (2013): "Blind Stealing: Experience and Expertise in a Mixed-Strategy Poker Experiment," working paper, University of Technology Sydney.

Wagenaar, Willem A. (1972): "Generation of random sequences by human subjects: A critical survey of literature," *Psychological Bulletin*, 77(1), 65–72.

Walker, Mark and Wooders, John (2001): "Minimax play at Wimbledon," *American Economic Review*, 91(5), 1521–1538.

Wooders, John (2010): "Does experience teach? Professionals and minimax play in the lab," *Econometrica*, 78(3), 1143–1154.