Serial correlation in National Football League play calling and its effects on outcomes

Emara, Noha and Owens, David and Smith, John and Wilmer, Lisa

Rutgers University-Camden, Haverford College

26 September 2014

Serial correlation in National Football League play calling and its effects on outcomes*

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

November 12, 2015

Abstract

We investigate the strategic behavior of highly informed agents playing zero-sum games under highly incentivized conditions. We examine data from 3455 National Football League (NFL) games from the 2000 season through the 2012 season, and categorize each play as "rush" or a "pass." We find that the pass-rush mix exhibits negative serial correlation: play types alternate more frequently than an independent stochastic process. This is an exploitable strategy, and we find that this serial correlation negatively affects play efficacy. 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 an effort to be perceived as unpredictable.

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. John Smith thanks Biblioteca de Catalunya. 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.

¶Florida State University, Department of Economics

1 Introduction

In two-player, zero-sum games with a unique mixed strategy Nash equilibrium, 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 randomize.

Research has shown that people have difficulty detecting and producing random sequences of the sort required for the execution of minimax.[1] Further, laboratory evidence suggests that mixing often does not occur as predicted, particularly when the data is analyzed at the individual level.[2] However, it is possible that subjects lack sufficient incentives or experience.

Rather than study this question in the laboratory, we go to the field, literally, examining strategic decisions in the National Football League (NFL). We categorize each offensive play as either a "rush" or a "pass" and investigate whether the chosen sequence exhibits serial correlation. 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 coaches and players earn large salaries and are under intense pressure to win, as evidenced by their frequent employment terminations. They can also confer with other highly trained and incentivized professionals,[3] and make detailed plans prior to the game.

We find that, despite these incentives, expertise, and opportunities for consultation and planning, observed play calling exhibits significant negative serial correlation. We also find that, according to two measures, play efficacy is adversely affected by excessive switching between play types. Our results suggest that teams could benefit from more clustered play sequences, which switch play type less frequently.

[1] For instance, see Wagenaar (1972), Bar-Hillel, and Wagenaar (1991), Rabin (2002), and Oskarsson et al. (2009). Note that research finds that there are differences between the generation of such sequences in decision problems and strategic settings (Rapoport and Budescu, 1992; Budescu and Rapoport, 1994).

[2] 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. (2014), Binmore, Swierzbinski, and Proulx (2001), Duffy, Owens, and Smith (2015), Geng et al. (2015), Levitt, List, and Reiley (2010), 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 (2015).

[3] 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.

While the existing literature can explain the negative serial correlation, it does not explain the reduced efficacy of plays associated with the negative serial correlation. Our results are consistent with defenses expecting the negative serial correlation exhibited by the offenses. This seems to be the case as clustered plays are more effective. Why would offenses employ an exploitable strategy that exhibits negative serial correlation? It is possible that they 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 by people who have difficulty detecting an independent sequence, and that the teams 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[4] long rectangular field. Two competing teams 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*, 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.[5]

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*, in order to advance the ball a minimum 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 a new set of four downs to advance the ball another ten yards. If the offensive team fails to net ten yards in the set of four plays,

[4] One yard is the equivalent of 0.9144 meters.
[5] A field goal is more likely to succeed when it is attempted closer to the opponent's end zone.

possession is transferred 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 possesses 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. On each play, teams benefit from accurately predicting the strategy of their opponent,[7] and by maintaining unpredictability in their own strategies. The space of available actions for both teams on each play is large and difficult to characterize. Fortunately, the strategy of the offensive team on each play can be meaningfully separated into one of two distinct categories: a *pass* or a *rush*.[8] We explore whether the pass-rush mix exhibits serial correlation.[9]

Rule violations are referred to as penalties, which 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.[10] For our purposes, the distinction is important because we can observe the type of play called on a live ball penalty, but not on a dead ball penalty.

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*.

---

[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 thereby rendering 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] Teams go to great lengths to obscure their strategies, and at times bend the rules to decipher those of their opponents. In 2007, the New England Patriots incurred one of the steepest punishments in NFL history for videotaping opposing coaches in order to learn their strategies.

[8] A pass is a play in which one player, normally the quarterback, attempts to advance the ball by throwing it forward to another player. A rush is a play in which a player attempts to advance the ball by carrying it.

[9] 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. For instance, see Bundrick and McGarrity (2014). Our data set does not allow us to distinguish between these possibilities. Therefore, we simply regard the decision making unit as the team.

[10] 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.

1.2 Related Literature

Laboratory studies find mixed evidence as to whether subjects play mixed strategies according to minimax. However, laboratory subjects may face relatively small material incentives and lack the necessary experience. In response to this critique, a growing literature examines mixed strategies in professional sports. These settings are characterized by large incentives and participants with a great deal of experience.[11]

Walker and Wooders (2001) examine the direction of serves in professional tennis matches. They find that the probability of success for serves to the right and serves to the left are not different, which is consistent with the equilibrium predictions. 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) perform 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. They largely find that participants mix according to equilibrium predictions (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 penalty kicks, 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.[12]

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).

[11] 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.

[12] Another line of research investigates whether the ability to mix according to the equilibrium predictions in a familiar strategic setting in the field translates to the ability to mix properly in an unfamiliar strategic setting in the laboratory. The conclusions in this literature are not uncontroversial (see Levitt, List, and Reiley, 2010; Palacios-Huerta and Volij, 2008; Van Essen and Wooders, 2015; Wooders, 2010).

McGarrity and Linnen (2010) examine play calling while restricting attention to first downs with a distance of ten yards. The authors analyze 11 NFL teams in the 2006 season and perform a test of runs for serial independence. They reject serial independence for only one of the 11 teams: their analysis supports the claim that play calling largely does not exhibit serial correlation. By contrast, analysis of our more extensive data finds serial correlation in play calling and that it leads to plays with reduced efficacy.

As we do, Kovash and Levitt (2009) find negative serial correlation across plays. Our study differs in that we employ different measures of efficacy. The authors 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 that 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 the equilibrium predictions. Our investigation does not compare the differences in outcomes between a pass and a rush, but explores whether play efficacy is affected by previous outcomes.

2 Data

Data, obtained from http://armchairanalysis.com for a small fee, includes an observation for each play 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 on football data,[13] we exclude data near the end of play and those characterized by a large score differential. Therefore, we omit from our analysis plays that occurred in the last 2 minutes of the second quarter, plays that occurred in the fourth quarter,[14] and plays that

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

[14] 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.

occurred when the absolute value of the point difference was 22 points or greater. After also excluding plays not from scrimmage (kickoffs, 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, which was included in the original data set, is whether the play is *successful* 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. Finally, we define a play as a *failure* if one yard or less is gained.

As this paper explores serial correlation in play calling, the assignment of a *previous play* to each play is crucial to the analysis. Plays that begin a possession are not assigned a previous play. Many complications in the assignment of a previous play arise due to the occurrence of penalties. Dead ball penalties are excluded from our analysis as the play type is not observed. The play following any sequence of dead ball penalties is assigned a previous play identical to the play type of the play preceding the sequence. A total of $209,963$ plays from scrimmage are assigned a previous play. We explore alternate methods of identifying a play's previous play, which we include in the supplemental online Appendix A. Our conclusions are robust to these alternate specifications.

3 Results

3.1 Summary statistics

In the analysis below, our independent variables include the down, the distance, the field position,[15] the point difference,[16] 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 1.

Table 1 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 than the game-level. The data includes 257,782 observations. We note that these calculations include plays that do not have a *previous* play.

Table 2 summarizes two measures of play efficacy: yards gained by the play and whether the play was successful.

[15] 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.

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

Table 2 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 explore the first of our primary research questions, whether the pass-rush mix exhibits serial correlation. Our dependent variable takes a value of 1 if the play is a pass and a 0 otherwise. We employ logistic regressions to determine which factor influences this outcome. Independent variables include the down, the distance, the field 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. We also control for various observables, such as whether the game was played in excessively cold conditions,[18] excessively windy conditions,[19] wet

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

[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.

conditions,[20] whether the game was played on grass,[21] and whether the offense was also the home team. Finally, we include specifications that account for the team-season fixed-effects.[22] We summarize this analysis in Table 3.

Table 3 Logistic regressions of serial correlation: Play is a pass

	(1)	(2)	(3)	(4)
Previous pass	-0.529^{***}	-0.546^{***}	-0.410^{***}	-0.425^{***}
	(0.0101)	(0.0101)	(0.0123)	(0.0123)
Previous failure	–	–	0.210^{***}	0.210^{***}
			(0.0177)	(0.0178)
Previous pass * Previous failure	–	–	-0.395^{***}	-0.399^{***}
			(0.0219)	(0.0220)
Fixed-effects?	*No*	*Yes*	*No*	*Yes*
-2 Log L	252380.07	251468.39	252048.58	251132.76
LR χ^2	35673.79^{***}	36585.47^{***}	36005.28^{***}	36921.10^{***}

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. Each regression has 209,963 observations. Note that * indicates significance at $p < 0.1$, ** indicates significance at $p < 0.01$, and *** indicates significance at $p < 0.001$.

First, the previous pass variable is negative and significant in each specification. This provides evidence that, even after controlling for down, distance, field position, and other observables, play calling exhibits significant negative serial correlation.[23] The Previous pass-Previous failure interaction estimate is negative and significant in both of the specifications where it appears, suggesting that play calling becomes even more negatively serially correlated following a failed play.[24]

[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 that also considers the playing surface. Unlike these authors, we do not find a significant effect.

[22] In our sample, there are 414 team-seasons. In the first two years of our sample there were 31 teams, as the Houston Texans did not yet exist. For the remaining years, there were 32 teams in the league.

[23] We run the regression (1) in Table 3, restricted to each of the 414 team-seasons. We find that the 2003 San Francisco 49ers have the smallest Previous pass estimate of -1.77 and the 2004 Saint Louis Rams have the largest at 0.37. See the supplemental online Appendix D for a discussion of the team-specific details and the exploration of the distribution of serial correlation across teams.

[24] We conduct the analysis similar to regression (4) in Table 3, on all four specifications of the previous play. Table $A3$ summarizes this analysis. We note that the specification of the previous play does not affect the results.

3.3 Serial correlation and yards gained

We now explore how serially correlated play calling affects play efficacy. As we do not observe 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 by analyzing the effect of serial correlation on yards gained. As independent variables, 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. We include team-seasons fixed-effects and the same control variables that were used in the analysis summarized in Table 3. Results of this regression are summarized in the (Prev 1) specification in Table $B1$ in the online appendix. We use the coefficients from this regression to estimate the differences in yards gained for both rushes and passes based on observable histories. We perform a Wald Test in order to test whether these differences are significant. This analysis is summarized in Table 4.

Table 4 Estimates of the difference in yards gained between a...

Rush following a rush and a rush following a pass	0.138*
	(0.0674)
Pass following a pass and a pass following a rush	0.211***
	(0.0511)
Rush following a failed rush and a rush following a failed pass	0.238*
	(0.118)
Rush following a non-failed rush and a rush following a non-failed pass	0.0368
	(0.0646)
Pass following a failed pass and a pass following a failed rush	0.0834
	(0.0776)
Pass following a non-failed pass and a pass following a non-failed rush	0.338***
	(0.0660)
R^2	0.05
F-value	24.89***

These estimates are based on 209,963 observations. These comparisons are based on the analysis summarized in Table $B1$. 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 4, suggesting that running the same play type as the previous play does not present a disadvantage in terms of yards gained. The first two rows only control for the play type and not whether the previous play was a failure. We find that a rush following a rush gains 0.14 more yards than a rush following a pass. We also find that a pass following a pass gains 0.21 more yards than a pass following a rush. Estimates are more pronounced when we also control for whether the previous play was a failure. We find that a rush gains 0.24 more yards more following a failed rush than following a failed pass. Also, a pass gains 0.34 more yards following a non-failed pass than following a non-failed rush. In summary, we find evidence that the efficacy of a play, as measured by yards gained, increases if it follows a play of the same type.[25]

3.4 Serial correlation and success

We find that negative serial correlation adversely affects the yards gained by a play. In this subsection, we employ an alternate measure of efficacy, the binary measure of success. This measure is different than yards gained, as passes obtain a larger average yards gained but rushes are more likely to be considered a success. We estimate differences, similar to that summarized in Table 4, but with this binary measure. We run the logistic regression that is summarized in the (Prev 1) specification in Table $B2$ in the online appendix. Given this analysis, we estimate the differences that we are interested in and conduct Wald tests on these estimates. We summarize these estimates in Table 5.

[25] In Table $A4$, we perform the analogous analysis on all four specifications of previous and find that the results are unchanged.

Table 5 Difference in success estimates between a...

Rush following a rush and a rush following a pass	0.123***
	(0.0176)
Pass following a pass and a pass following a rush	0.0274*
	(0.0131)
Rush following a failed rush and a rush following a failed pass	0.251***
	(0.0313)
Rush following a non-failed rush and a rush following a non-failed pass	−0.0055
	(0.0160)
Pass following a failed pass and a pass following a failed rush	0.0110
	(0.0202)
Pass following a non-failed pass and a pass following a non-failed rush	0.0437**
	(0.0165)
-2 Log L	269807.43

These estimates are based on $209,963$ observations. These comparisons are based on the analysis summarized in Table $B2$. Based on the Wald test of the estimates, * indicates significance at $p < 0.1$, ** indicates significance at $p < 0.01$, and *** indicates significance at $p < 0.001$.

There are no significant and negative estimates in Table 5, again suggesting that plays are more successful when they follow plays of the same type. In addition, our results again 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. We also 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.[26] The analysis summarized in Table 5 reinforces that of Table 4, and provides evidence that play efficacy is 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 differs sufficiently across downs, so we analyze second and third downs separately. We begin by restricting our attention to second down and conduct an analysis identical to that summarized in Table 3. We summarize this analysis in Table 6.

[26] We also conduct the estimates as summarized in Table 5 but with an alternate measure of success. This can be found in Table $C1$, in the online appendix.

Table 6 Logistic regressions of serial correlation: Play is a pass on second down

	(1)	(2)	(3)	(4)
Previous pass	-0.953^{***}	-0.973^{***}	-0.639^{***}	-0.655^{***}
	(0.0157)	(0.0159)	(0.0226)	(0.0228)
Previous failure	–	–	0.203^{***}	0.204^{***}
			(0.0231)	(0.0233)
Previous pass * Previous failure	–	–	-0.622^{***}	-0.629^{***}
			(0.0319)	(0.0322)
Fixed-effects?	*No*	*Yes*	*No*	*Yes*
-2 Log L	109300.05	108263.19	108919.52	107882.13
LR χ^2	10665.72^{***}	11702.59^{***}	11046.26^{***}	12083.64^{***}

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. Each regression has $86,645$ observations. Note that * indicates significance at $p < 0.1$, ** indicates significance at $p < 0.01$, and *** indicates significance at $p < 0.001$.

We again find evidence of negative serial correlation. We also find that a second consecutive pass on second down is even less likely after a *failed* pass on first down.

Next we investigate whether the negative serial correlation on second down affects the efficacy of plays. We conduct an analysis similar to that summarized in Table 4, restricting attention to second down. This analysis is summarized in Table $B3$ in the online appendix. From this analysis we estimate the differences, which are presented in Table 7.

Table 7 Estimates of the difference in yards gained on second down between a...

Rush following a rush and a rush following a pass	0.239^{**}
	(0.0857)
Pass following a pass and a pass following a rush	0.452^{***}
	(0.0851)
Rush following a failed rush and a rush following a failed pass	0.389^{**}
	(0.1328)
Rush following a non-failed rush and a rush following a non-failed pass	0.0886
	(0.107)
Pass following a failed pass and a pass following a failed rush	0.315^{**}
	(0.108)
Pass following a non-failed pass and a pass following a non-failed rush	0.589^{***}
	(0.131)
R^2	0.05
F-value	10.38^{***}

These estimates are based on $86,645$ observations. These comparisons are based on the analysis summarized in Table $B3$. 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 again find no significant and negative estimate of the difference in yards gained on second down. In fact, the differences on second down appear more pronounced. While Table 4 only has 2 of the 6 conditions that are significant at 0.01, Table 7 has 5 of the 6. As in the previous analyses, the adverse impact of switching play type becomes more pronounced when the previous play was a failure.[27] 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.

3.6 Analysis of third down

Table 8 shows the results of the analysis conducted in the same manner as that summarized in Table 6, with attention restricted to third down.

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

	(1)	(2)	(3)	(4)
Previous pass	-0.0293	-0.0281	-0.0603^{*}	-0.0593^{*}
	(0.0227)	(0.0231)	(0.0302)	(0.0307)
Previous failure	–	–	0.139^{***}	0.144^{***}
			(0.0320)	(0.0324)
Previous pass * Previous failure	–	–	0.0314	0.0306
			(0.0460)	(0.0467)
Fixed-effects?	*No*	*Yes*	*No*	*Yes*
-2 Log L	50970.84	50245.27	50931.61	50204.85
LR χ^2	5882.85^{***}	6608.41^{***}	5922.07^{***}	6648.84^{***}

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. Each regression has $54,922$ observations. Note that * indicates significance at $p < 0.1$, ** indicates significance at $p < 0.01$, and *** indicates significance at $p < 0.001$.

Evidence of negative serial correlation on third down plays is weaker and the estimates are significant only at the 0.1 level. In the specifications that do not include Previous failure, we

[27] We also conduct the analogous analysis consisting of the success measure, rather than the yards gained measure, on second down plays. This is available from the corresponding author upon request.

do not find evidence of serial correlation, and the Previous pass-Previous failure interaction is not significant in either of our specifications.

Table 9 shows the results of our analysis, similar to that in the previous subsection, restricting attention to third down plays. The analysis is summarized in Table $B4$, in the online appendix. From this analysis we estimate the relevant differences, which are presented in Table 9.

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

Rush following a rush and a rush following a pass	-0.101
	(0.180)
Pass following a pass and a pass following a rush	0.0331
	(0.0940)
Rush following a failed rush and a rush following a failed pass	-0.324
	(0.279)
Rush following a non-failed rush and a rush following a non-failed pass	0.122
	(0.227)
Pass following a failed pass and a pass following a failed rush	-0.141
	(0.124)
Pass following a non-failed pass and a pass following a non-failed rush	0.207
	(0.140)
R^2	0.05
F-value	7.17^{***}

These estimates are based on $54,922$ observations. These comparisons are based on the analysis summarized in Table $B4$. Based on the Wald test of the estimates, * indicates significance at $p < 0.1$, ** indicates significance at $p < 0.01$, and *** indicates significance at $p < 0.001$.

Table 9 does not contain a single significant relationship, suggesting that the weaker evidence of negative serial correlation found in Table 8 does not 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 9.[28]

[28] 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.

4 Conclusion

We use play-level data to 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, and find that the play calling exhibits negative serial correlation. In other words, play type switches more than an independent stochastic process.

We also find that play efficacy is affected by previous actions and previous outcomes. In particular, we find that rush plays following rush plays earn more yards and are more likely to be successful than rush plays following pass plays. We find the analogous result for pass plays. Given that the yards and success measures exhibit 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 the adverse impact of excessive switching between play types becomes more pronounced when the previous play was a failure. We also conduct an analysis of serial correlation separately for second and third downs. Second down plays exhibit stronger negative serial correlation and the adverse effects of serial correlation are more pronounced than on third down.

What are we to make of our results? How could it be the case that experienced, incentivized decision makers who can consult other professionals, and can make detailed plans, mix in a manner that is exploitable?

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 may explain 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.[29] 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 mixing randomly. 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.[30] 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.

Perhaps teams feel pressure not to repeat the play type on offense, in order to avoid criticism for being too "predictable" by fans, media, or executives 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.[31] 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 that result from the negative serial correlation seems to be the explanation most consistent with our data.

We stop short of claiming that the teams are not acting according to the minimax predictions. While we use two reasonable measures of efficacy, the ultimate objective of the teams is to win the game. Although success on any given play is related to winning the game, it is not obvious that an equilibrium within each particular play is identical to the equilibrium in the extensive form of the game. For instance, Walker, Wooders, and Amir (2011) study a setting in which two agents engage in a sequence of plays, each with one of two possible outcomes. This continues until a winner of the overall game is declared. The authors show the conditions under which the equilibrium within any particular play is also the equilibrium of the larger overall game. As our game is more complicated than theirs, it is not obvious

[29] See Wagenaar (1972), Bar-Hillel, and Wagenaar (1991), Rabin (2002), and Oskarsson et al. (2009).

[30] For instance, see Halpern and Pass (2015).

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

that this result holds in our setting. Further, providing such an argument is beyond the scope of this paper. On the other hand, we suspect that a formal argument will show that, given reasonable restrictions on the preferences of the teams and the setting, the exhibition of serial correlation would not be consistent with the minimax of the overall game. We leave this issue for future work.

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

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

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

Batzilis, Dimitris, Sonia Jaffe, Steven Levitt, John A. List, and Jeffrey Picel, 2014. "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, Joe Swierzbinski, and Chris Proulx, 2001. "Does minimax work? An experimental study." *Economic Journal* 111(473), 445–464.

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

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

Bundrick, Jacob and Joseph P. McGarrity, 2014. "Strategic Play in the NFL's Offensive Play Calling." *International Advances in Economic Research* 20(3), 339–340.

Buzzacchi, Luigi and Stefano Pedrini, 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., Steven Levitt, and Timothy 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, Germán, 2007. "Penalty Kicks in Soccer An Alternative Methodology for Testing Mixed-Strategy Equilibria." *Journal of Sports Economics* 8(5), 530–545.

Duffy, Sean, David Owens, and John Smith, 2015. "Cognitive load and mixed strategies." working paper, Rutgers University-Camden.

Geng, Sen, Yujia Peng, Jason Shachat, and Huizhen Zhong, 2015. "Adolescents, cognitive ability, and minimax play." *Economics Letters* 128, 54–58.

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

Halpern, Joseph Y. and Rafael Pass, 2015. "Algorithmic rationality: Game theory with costly computation." *Journal of Economic Theory* 156, 246–268.

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

Kovash, Kenneth and Steven D. Levitt, 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., John A. List, and David 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.

McGarrity, Joseph P. and Brian Linnen, 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 Barry Sopher, 1997. "Learning and decision costs in experimental constant sum games." *Games and Economic Behavior* 19(1), 97–132.

Mookherjee, Dilip and Barry Sopher, 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., Leaf Van Boven, Gary H. McClelland, and Reid Hastie, 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 Oscar Volij, 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 Wilfred Amaldoss, 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 Wilfred Amaldoss, 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 Richard B. Boebel, 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 David V. Budescu, 1992. "Generation of random series in two-person strictly competitive games." *Journal of Experimental Psychology: General* 121(3), 352–363.

Reed, Derek D., Thomas S. Critchfield, and Brian K. Martens, 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., Jason Shachat, and Mark Walker, 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 J. Todd Swarthout, 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 Thomas S. Critchfield, 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 John Wooders, 2015. "Blind stealing: Experience and expertise in a mixed-strategy poker experiment." *Games and Economic Behavior* 91, 186–206.

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

Walker, Mark, John Wooders, and Rabah Amir, 2011. "Equilibrium play in matches: Binary Markov games." Games and Economic Behavior 71(2), 487–502.

Walker, Mark and John Wooders, 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.

Supplemental Online Appendix

In Appendix A, we explore the robustness of our specification of the previous play. In Appendix B, we present the regressions that formed the basis of the estimates in Tables 4, 5, 7, and 9. In Appendix C, we employ an alternate specification of the success of the play. Finally in Appendix D, we include an exploration of the distribution of the serial correlation across teams.

Appendix A: Different specifications of the previous play

In the body of the paper we describe and analyze a particular specification of the previous play. In this appendix, we refer to this as *Previous 1*. In order to learn whether our results are robust to different ways of assigning a previous play, we describe and analyze three alternate specifications. One alternate specification is to assign the previous play as in Previous 1, except that plays following dead ball penalties are not assigned previous plays. Unlike Previous 1, here the teams do not consider the information prior to a dead ball penalty. We refer to this specification as *Previous 2*. There are $203,791$ plays from scrimmage with an observation involving Previous 2.

Another way to assign 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, the play following any sequence of penalties is assigned a previous play identical to the play type of the play preceding the sequence. We refer to this specification as *Previous 3*. There are $202,329$ plays from scrimmage with an observation involving Previous 3.

Finally, one could assign the previous play as in Previous 3, except that plays following live ball or dead ball penalties are not assigned previous plays. In other words, here the teams do not consider the information prior to any penalty. We refer to 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 $A1$, where a check indicates that it satisfies the criteria.

Table A1 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 among 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 is a rush. The fifth is a dead ball penalty. The sixth and seventh plays are passes. Table $A2$ illustrates the differences among the previous classifications in this example.

Table A2 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.

Table $A3$ reproduces the analysis in regression (4) of Table 3 for all four specifications of the previous play. Note that the (Prev 1) regression in Table $A3$, is identical to regression (4) in Table 3.

Table A3 Logistic regressions of serial correlation: Play is a pass

	(Prev 1)	(Prev 2)	(Prev 3)	(Prev 4)
Previous pass	-0.425^{***}	-0.423^{***}	-0.410^{***}	-0.420^{***}
	(0.0123)	(0.0125)	(0.0125)	(0.0128)
Previous failure	0.210^{***}	0.196^{***}	0.179^{***}	0.127^{***}
	(0.0178)	(0.0182)	(0.0179)	(0.0188)
Previous pass * Previous failure	-0.399^{***}	-0.426^{***}	-0.446^{***}	-0.479^{***}
	(0.0220)	(0.0224)	(0.0224)	(0.0230)
-2 Log L	251132.76	243932.73	241843.88	232353.73
LR χ^2	36921.10^{***}	35952.74^{***}	35954.28^{***}	35188.28^{***}
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 $A4$ reproduces the estimates of Table 4 for all four specifications of the previous play. Note that the (Prev 1) estimates Table $A4$ are identical to estimates in Table 4.

Table A4 Estimates of the difference in yards gained between a...

	(Prev 1)	(Prev 2)	(Prev 3)	(Prev 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 $B1$. Based on the Wald test of the estimates, * indicates significance at $p < 0.1$, ** indicates significance at $p < 0.01$, and *** indicates significance at $p < 0.001$.

Table $A5$ reproduces the estimates of Table 5 for all four specifications of the previous play. Note that the (Prev 1) estimates Table $A5$ are identical to estimates in Table 5.

Table A5 Difference in success estimates between a...

	(Prev 1)	(Prev 2)	(Prev 3)	(Prev 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 $B2$. Based on the Wald test of the estimates, * indicates significance at $p < 0.1$, ** indicates significance at $p < 0.01$, and *** indicates significance at $p < 0.001$.

Table $A6$ reproduces the analysis in regression (4) of Table 6 for all four specifications of the previous play. Note that the (Prev 1) regression in Table $A6$, is identical to regression (4) in Table 6.

Table A6 Logistic regressions of serial correlation: Play is a pass on second down

	(Prev 1)	(Prev 2)	(Prev 3)	(Prev 4)
Previous pass	−0.655***	−0.661***	−0.648***	−0.655***
	(0.0228)	(0.0231)	(0.0229)	(0.0233)
Previous failure	0.204***	0.205***	0.226***	0.194***
	(0.0233)	(0.0239)	(0.0237)	(0.0245)
Previous pass * Previous failure	−0.629***	−0.645***	−0.675***	−0.703***
	(0.0322)	(0.0327)	(0.0327)	(0.0332)
-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. The (Prev 1) specification in Table $A6$, is identical to specification (4) in Table 6. Note that * indicates significance at $p < 0.1$, ** indicates significance at $p < 0.01$, and *** indicates significance at $p < 0.001$.

Table $A7$ reproduces the estimates of Table 7 for all four specifications of the previous play. Note that the (Prev 1) estimates Table $A7$ are identical to estimates in Table 7.

Table A7 Estimates of the difference in yards gained on second down between a

	(Prev 1)	(Prev 2)	(Prev 3)	(Prev 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 $B3$. Based on the Wald test of the estimates, * indicates significance at $p < 0.1$, ** indicates significance at $p < 0.01$, and *** indicates significance at $p < 0.001$.

Table $A8$ reproduces the analysis in regression (4) of Table 8 for all four specifications of the previous play. Note that the (Prev 1) regression in Table $A8$, is identical to regression (4) in Table 8.

Table A8 Logistic regressions of serial correlation: Play is a pass on third down

	(Prev 1)	(Prev 2)	(Prev 3)	(Prev 4)
Previous pass	−0.0593*	−0.0549*	−0.0529*	−0.0514
	(0.0307)	(0.0312)	(0.0311)	(0.0313)
Previous failure	0.144***	0.139***	0.142***	0.130***
	(0.0324)	(0.0332)	(0.0331)	(0.0334)
Previous pass * Previous failure	0.0306	−0.00764	0.000378	−0.0182
	(0.0467)	(0.0478)	(0.0479)	(0.0485)
-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. The (Prev 1) specification in Table $A8$, is identical to specification (4) in Table 8. Note that * indicates significance at $p < 0.1$, ** indicates significance at $p < 0.01$, and *** indicates significance at $p < 0.001$.

Table $A9$ reproduces the estimates of Table 9 for all four specifications of the previous play. Note that the (Prev 1) estimates Table $A9$ are identical to estimates in Table 9.

Table A9 Estimates of the difference in yards gained on third down between a

	(Prev 1)	(Prev 2)	(Prev 3)	(Prev 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 $B4$. Based on the Wald test of the estimates, * indicates significance at $p < 0.1$, ** indicates significance at $p < 0.01$, and *** indicates significance at $p < 0.001$.

Appendix B: Summary of regressions

Table $B1$ presents a summary of the regressions from the estimates presented in Tables 4 and $A4$ for all four definitions of the previous play

Table B1 Regressions of yards gained

	(Prev 1)	(Prev 2)	(Prev 3)	(Prev 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 $p < 0.001$.

We now explain how the differences in Table 4 are constructed from the estimates in Table $B1$. The difference between a Rush following a rush and a rush following a pass and the difference between a Pass following a pass and a pass following a rush are estimated from a model that does not control for the Previous failure variable. Therefore these differences do not obviously follow from the analysis summarized in Table $B1$. However, the other differences follow from the analysis summarized in Table $B1$. The difference between a Rush following a non-failed rush and a rush following a non-failed pass is equal to the Same play type as previous estimate, 0.0368. The difference between a Rush following a failed rush and a rush following a failed pass is equal to the estimate for Same (0.0368) plus the estimate for the Previous failure-Same interaction (0.202). The difference between a Pass following a non-failed pass and a pass following a non-failed rush is equal to the sum of the estimates for Same (0.0368) and the Pass-Same interaction (0.301). The difference between a Pass following a failed pass and a pass following a failed rush is equal to the sum of the estimates for Same (0.0368), the Pass-Same interaction (0.301), the Pass-Previous failure-Same interaction (−0.456), and the Previous failure-Same interaction (0.202).

Table $B2$ presents a summary of the regressions from the estimates presented in Tables 5 and $A5$ for all four definitions of the previous play

Table B2 Logistic regressions of a successful play

	(Prev 1)	(Prev 2)	(Prev 3)	(Prev 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 $B3$ presents a summary of the regressions from the estimates presented in Tables 7 and $A7$ for all four definitions of the previous play

Table B3 Regressions of yards gained on second down

	(Prev 1)	(Prev 2)	(Prev 3)	(Prev 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 $B4$ presents a summary of the regressions from the estimates presented in Tables 9 and $A9$ for all four definitions of the previous play

Table B4 Regressions of yards gained on third down

	(Prev 1)	(Prev 2)	(Prev 3)	(Prev 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 $p < 0.001$.

Appendix C: Alternate specification of success

Here we perform the estimates as conducted in Table 5, however, we use a different measure of success. This alternate measure of success uses the same criteria for first, third, and fourth downs as the original measure. However, in order for the play to be a considered success on second down, 70% of the distance, rather than 60%, must be gained.

Table C1 Difference in alternate success estimates between a...

Rush following a rush and a rush following a pass	0.107^{***}
	(0.0181)
Pass following a pass and a pass following a rush	0.00583
	(0.0132)
Rush following a failed rush and a rush following a failed pass	0.248^{***}
	(0.0325)
Rush following a non-failed rush and a rush following a non-failed pass	-0.0328^{*}
	(0.0160)
Pass following a failed pass and a pass following a failed rush	-0.0006
	(0.0205)
Pass following a non-failed pass and a pass following a non-failed rush	0.0122
	(0.0164)
-2 Log L	267339.89

These estimates are based on $209,963$ observations. Based on the Wald test of the estimates, * indicates significance at $p < 0.1$, ** indicates significance at $p < 0.01$, and *** indicates significance at $p < 0.001$.

In both Table 5 and Table $C1$, we see that a rush following a rush is significantly more likely to be successful than a rush following a pass. We also see that a rush following a failed rush is significantly more likely to be successful than a rush following a failed pass. However, we note that there are differences between the estimates presented in Table 5 and those in Table $C1$. We see that the negative effects of the serial correlation are less severe for pass plays. We also see that there is a negative significant estimate for the difference between a rush followed by a non-failed rush and a rush followed by a non-failed pass play.

Appendix D: Team-specific serial correlation

We now investigate the team-specific aspects of the serial correlation in play calling. We conduct the analysis of regression (3) in Table 3 restricted to each of the 414 team-seasons in our data set. For every team-season we note the estimate of the Previous pass coefficient. We then compare these estimates to other characteristics of the teams. Perhaps teams that are more successful exhibit less serial correlation? We therefore compare the estimates to the number of regular season wins during that season.[32] The Wins variable has a minimum of 0 and a maximum of 16. It is also possible that head coaching positions that are less stable are more likely to exhibit serial correlation. We therefore compare the estimates to the number of head coaches employed by that particular team during our sample period. This Coaches variable ranges from 1 (Philadelphia Eagles) to 8 (Oakland Raiders). Finally, it is possible that the longer a head coach is employed, the less pressure there will be to appear to be "unpredictable." Therefore we define the Tenure length variable, which measures the length of the head coach's tenure. If there was a mid-season change in the head coach then Tenure length is 0. If a first-year head coach begins and ends the season as head coach, the Tenure

[32] There were three ties in our sample period. In the event of a tie, both teams were awarded 0.5 wins.

length is 1. If the head coach was not replaced before the end of the second year, the Tenure length is 2. This continues until either the end of the sample period or the replacement of the head coach. In Table $D1$ we run regressions with the team-season Previous pass estimates as the dependent variable and the independent variables mentioned above.

Table D1 Previous pass coefficient estimate by team-season

	(1)	(2)	(3)
Wins	-0.00176	–	–
	(0.00642)		
Coaches	–	0.00140	–
		(0.0104)	
Tenure length	–	–	-0.00226
			(0.00604)
R^2	0.00018	0.000044	0.00034

Each regression has 414 observations. Note that * indicates significance at $p < 0.1$, ** indicates significance at $p < 0.01$, and *** indicates significance at $p < 0.001$.

We do not find a relationship between the serial correlation estimates and either the team or team-season characteristics. Below, in Figure $D1$, we show the scatterplot for the number of average wins for a team and the average of the Previous pass estimates for the team.

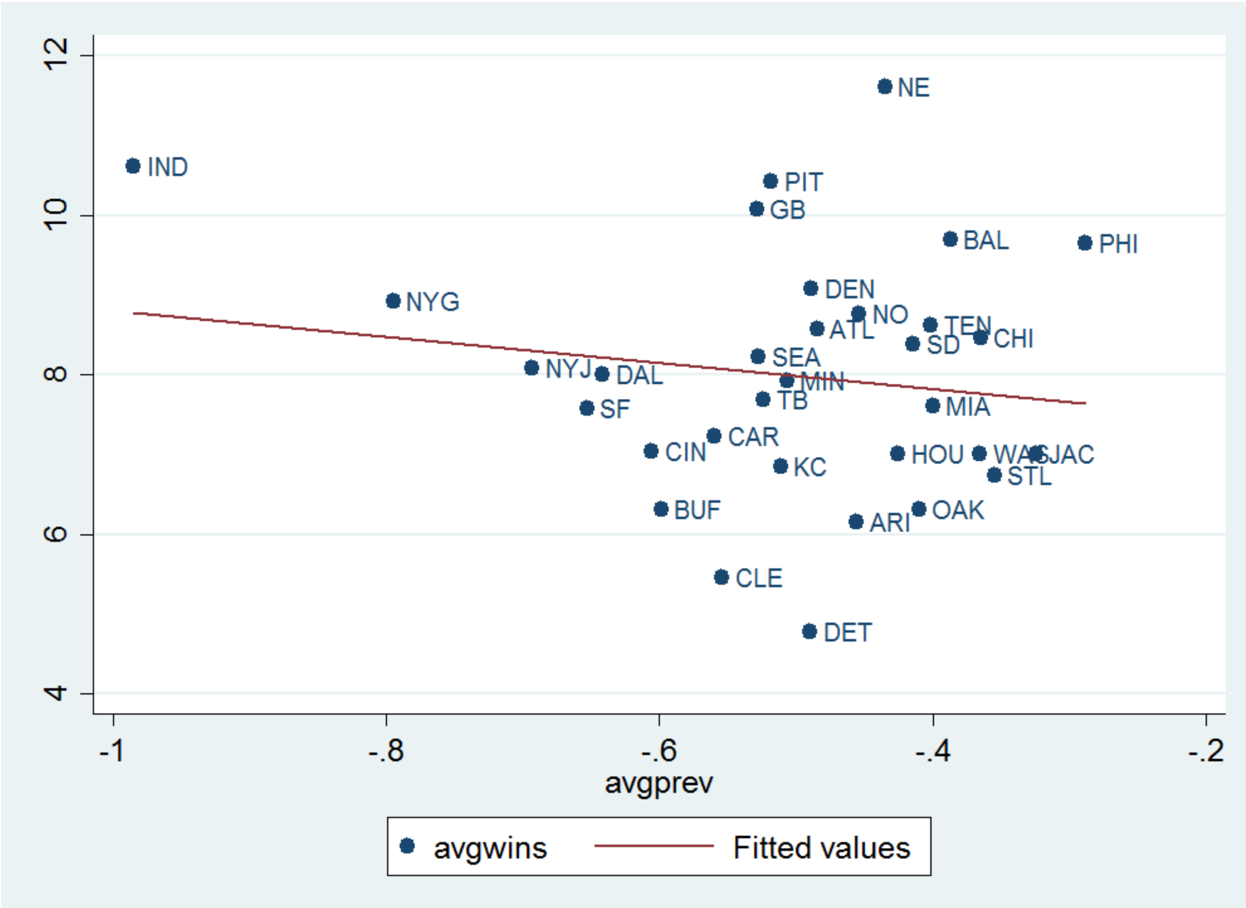
12
10
8
6
4
IND
NYG
NYJ
DAL
SF
CIN
CAR
BUF
CLE
DET
NE
PIT
GB
BAL
PHI
DEN
NO
ATL
TEN
SD
CHI
SEA
MIN
TB
MIA
KC
HOU
WAS
JAC
STL
ARI
OAK
-1
-.8
-.6
-.4
-.2
avgprev
avgwins
Fitted values

Figure D1