Southern Economic Journal 2010, 76(3), 791–810

Pass or Run: An Empirical Test of the Matching Pennies Game Using Data from the National Football League

Joseph P. McGarrity* and Brian Linnen†

This article examines play calling in the National Football League (NFL). It finds that a mixed strategy equilibrium game explains NFL play calling better than standard optimization techniques. When a quarterback is injured and replaced with a less capable backup, standard optimization theory suggests that the offense will run more often, passing less. Our game theoretic model predicts that the offense will not change its play calling because the defense will play against the run more often. Using every first half play from the 11 teams that had a starting quarterback miss action because of injury in the 2006 season, we find that the injury did not alter the likelihood that the offense would pass. We also find that coaches randomly mix passing and running plays, as the mixed strategy games predict.

JEL Classification: C72, C93

1. Introduction

Standard economic optimizing techniques, such as isoquant and isocost analysis, predict that when an input becomes less productive, firms will use it less and use more of their other inputs instead. Consider the case of American football. The offense has two main options. It can run the ball, or it can pass the ball. Next, suppose a starting National Football League (NFL) quarterback is injured. If we assume that the starting quarterback is more proficient than his replacement, the standard optimizing approach suggests that the coach will have the replacement quarterback throw less frequently. That is, the coach will decrease his use of the input with diminished productivity, and he will use the running back, whose productivity has not changed, more often.

However, the standard optimizing approach does not account for the fact that the offensive output is dependent, in part, upon decisions made by the defense. The defense tries to undermine the offense whenever it can. If the defense believes the offense will run the ball, the defense will try to move its players toward the line of scrimmage, thereby closing any holes the running back can run through. If the defense believes the offense will pass, once the ball is snapped, the defense might assign more players to rush the quarterback or assign more players to cover potential receivers. The game theory approach is able to deal with this adversarial

* Economics Department, 304 BBA, UCA, Conway, AR 72035, USA; E-mail joem@uca.edu; corresponding author.

† Werner Enterprises, 14507 Frontier Road, Omaha, NE 68138, USA; E-mail ml03001@gmail.com.

We would like to thank several people who made suggestions and comments that led to significant improvements in our article. We thank two anonymous referees, Tim Bisping, James Boone, Mike Casey, Craig Depken, Ling He, Fred Hickling, Bruce Johnson, Steven Levitt, Dan Sutter, and Robert Tollison. We also thank the participants of the 2007 International Atlantic Economic Conference.

Received January 2008; accepted March 2009.

relationship. Interestingly, in the case of the injured starting quarterback, the game theory prediction is much different than the isoquant/isocost prediction.

This article develops a game theoretic model that predicts that an offense will not change its play calling when its starting quarterback cannot play because of an injury. Instead, the defense will play against the run more often. Although the defensive play calling is not readily observable, it is very clear whether the offense calls a pass or a run. This article tests whether starting quarterbacks pass more often than their replacements. Using data from the 11 teams that replaced an injured quarterback in the 2006 NFL season, we find that an injury did not change offensive play calling.

The matching pennies game, the mixed strategy game developed in this article, is the standard theoretical approach that one would use to explain how actors behave if they benefit when other players cannot accurately guess their actions. Therefore, it is interesting that the matching pennies game has remained so important theoretically when the empirical evidence indicates that people do not act as the model suggests that they should. However, almost all of the empirical tests have been conducted in a laboratory setting.[1] Walker and Wooders (2001) argue that the laboratory may not be the best setting for testing the applicability of mixed strategy equilibria. They note that playing a game well is difficult, and laboratory participants may not have the incentive to learn to play the game well because their payoffs are often small and because they do not get a chance to keep playing the game after the experiment is over.

Walker and Wooders argue that sports may offer a better setting for testing whether people adopt mixed strategies in a matching pennies game.[2] Professional athletes invest heavily in learning their game and would very likely be able to take advantage of a player who did not adopt the optimal mix of strategies. They also argue that professional athletes have enough at stake so that they will learn the best strategic approaches to their game. To date, very few articles have tested the predictions of the matching pennies game outside the laboratory. We have been able to find only six such articles, and all of them use sports data to conduct the tests.

The aforementioned Walker and Wooders (2001) is the first article to use non-laboratory data to test the predictions of a matching pennies game. In support of the game theoretic predictions, they find that tennis players mixed their serves between an opponent's backhand and forehand, so that a serve to either side had an equal probability of being a winning point. However, they found that the service choice was not serially independent as the mixed strategy equilibrium requires. They report that serves alternated sides too often. In a comment Hsu, Huang, and Tang (2007) replicate these results with a larger sample and find evidence in support of both of these two game theoretic predictions. Klaassen and Magnus (2009) find that tennis players at Wimbledon do not adopt an optimal service strategy. They suggest that optimal strategies would increase the financial winnings of men by 18.7% and increase the financial winnings of women by 32.8%.

Chiappori, Levitt, and Groseclose (2002) find that soccer players mix the location of penalty kicks in a way consistent with the predictions of a matching pennies game. Using a larger sample of soccer penalty kicks, Palacios-Huerta (2003) reaches the same conclusion.

---

[1] See chapter 3 of Camerer (2003).

[2] Sports have been used as a setting to conduct tests of various economic theories. The authors have argued that the results can give us insight into how humans make choices in environments other than sports. For example, see McCormick and Tollison (1984), Goff, Shughart, and Tollison (1997), Goff, McCormick, and Tollison (2002), Heckelman and Yates (2003), and Bradbury and Drinen (2007). Also, sports has become such a large business that scholars are concerned with sports-related decisions considered by themselves—without claiming that the results are applicable outside of the realm of sports. See, for example, Sutter and Winkler (2003) and Poitras and Hadley (2006).

Coloma (2007) finds evidence of mixed strategy play when using the data of Chiappori, Levitt, and Groseclose (2002) and employing an empirical strategy of estimating a simultaneous equation model. The three articles on soccer penalty kicks all find that mixed strategies, which did not seem to explain much in the laboratory, do a very good job explaining real world behavior. Taken together, the articles that analyze tennis serves also find support for mixed strategy play; although, this support is weaker than that found in the articles using soccer data. Using a different sport (football), this article adds further evidence that mixed strategies are actually played outside of the laboratory.

One of the clear predictions of a mixed strategy equilibrium is that the expected utility of all actions should be the same. In the case of soccer penalty kicks, a goal is worth the same amount of points regardless of which side of the goal the ball enters. This implies that a kick to the right side of the goal should have an equal chance of scoring as a kick to the left side of the goal. In football we cannot assume that, in equilibrium, the expected yards from a pass yield the same number of yards as the expected yards from a run. We cannot, because a pass is riskier than a run, and the team's offense would have to expect more yards, on average, from a pass to be willing to accept the additional risk of a turnover. However, we can use comparative statics to determine whether offenses are using a rational mixed strategy. An exogenous event such as an injury to the starting quarterback makes different predictions about the new equilibrium depending upon how the team optimizes. We can differentiate whether the offense is using a mixing strategy or optimizing without considering possible defensive reaction.

2. Game Theoretic Model

In the football play calling game, the offense and defense are adversaries. The offense can call a passing play or a running play. The defense must decide to defend against the run, or defend against the pass. The offense will do well when (i) it calls a passing play and its opponents play a run defense or (ii) it calls a running play when the defense is set up to defend a pass. Similarly, the defense does well when it tries to defend against the type of play the offense calls. The offense and defense must simultaneously decide what play to call. They will try to disguise their calls so it is not apparent what play was called, even as the play is beginning to be executed. For example, in a play action pass, a quarterback fakes a hand-off, and it is not apparent that the play is a pass until after the fake.

The game is characterized in Table 1. The offense's payoffs for each outcome are given by the capital letters (A, B, C, D). Because football is a zero sum game, and every yard the offense gains the defense gives up, the defense's payoffs are represented by the negative of the offense's payoffs ($-A$, $-B$, $-C$, $-D$). These letters do not represent specific cardinal utilities for each outcome. Instead, this setup is much more general. The letters are variables, and any level of utility can be plugged in for a letter.

Consider the four outcomes. The outcomes in the chart's northwest corner (A, $-A$) and the chart's southeast corner (D, $-D$) represent the times that the defense guesses correctly and defends against the type of play that the offense actually called. The outcomes in the northeast corner (B, $-B$) and those in the southwest corner (C, $-C$) represent the times that the offense fooled the defense and can run a play against a defense that is not designed to stop that type of play.

Table 1. Football Play Calling Game with Starting Quarterback

	Defensive Play	
Offensive Play	Defend Pass (y)	Defend Run $(1-y)$
Pass (a)	$A, -A$	$B, -B$
Run $(1-a)$	$C, -C$	$D, -D$

Payoffs: offense, defense.

We can say something about the offense's preference orderings. If the defense will attempt to stop the run, the offense would be better off if it passed rather than ran the ball, so we know that $B > D$. If the defense attempts to stop the pass, the offense would rather run than pass, so $C > A$. Given these preferences, the offense does not have a dominant pure strategy. Its best strategy depends upon what it believes the defense will do.[3]

A similar story can be told for the defense. Its best strategy depends upon what the offense does. If the offense wants to pass the ball, the defense should try and stop the pass. Therefore, we know that $-A > -B$. In the other case, when the offense wants to run the ball, the defense is better off if it tries to stop the run, so $-D > -C$. This game is a matching pennies game. One player wants to pick the same option as the opponent, but the other player wants to take a different option. As with a typical matching pennies game, there is no pure strategy Nash equilibrium.

In any given box, one player will want to move counter-clockwise to the next box. However, there is a mixed strategy Nash equilibrium. Suppose the offense passes the ball with a probability of a, and the defense defends against the pass with a probability of y. We can use the payoff-equating method to solve for both probabilities. First, consider the offense's payoffs. In the mixed strategy equilibrium, the offense will be indifferent between its two options; that is, the offense's expected utility from passing and running must be identical. The payoff-equating method exploits this equality to solve for the unknown probabilities. The offense's payoffs from its two strategies are

$$\text{Payoff}_{(\text{pass})} = Ay + (B(1-y)),$$
$$\text{Payoff}_{(\text{run})} = Cy + (D(1-y)).$$

By setting both payoffs equal we get

$$Ay + (B(1-y)) = [Cy + D(1-y)].$$

This can be simplified to

$$Ay + B - By = Cy + D - Dy,$$
$$Ay - By - Cy + Dy = D - B,$$
$$y(A - B - C + D) = D - B,$$

so

$$y = D - B/[(D-B) + (A-C)]. \tag{1}$$

[3] A glance at Table 6 will reveal that no team passes or runs all the time. One can imagine that run may be a dominant pure strategy (regardless of whether the defense plays against the run or pass) in high school football. However, this is obviously not the case in pro football.

Therefore, the probability that the defense will defend against the pass is solved by using the payoffs from the offense. By similar logic, the probability that the offense will pass, a, can be solved by referring to the defense's payoffs for (i) defending against the pass and (ii) defending against the run. The payoffs from each action are

$$\text{Payoffs}_{(\text{defend against pass})} = a(-A) + (1-a)(-C),$$
$$\text{Payoffs}_{(\text{defend against run})} = a(-B) + (1-a)(-D).$$

Because the defense is indifferent between the two strategies in equilibrium, we can set the payoffs equal:

$$a(-A) + (1-a)(-C) = a(-B) + (1-a)(-D).$$

This can be simplified to

$$-aA - C + aC = -aB - D + aD,$$
$$-aA + aC + aB - aD = C - D,$$
$$a(-A + C + B - D) = (C - D),$$

so

$$a = C - D/[(C-D) + (B-A)]. \tag{2}$$

Equations 1 and 2 give the general results for the probability that a defense will defend against the run, y, and the probability that the offense will throw a pass, a. The model and results that we derived in this section should be considered as a representative situation faced by the offense and defense. Of course, the values of the payoffs will vary with the game situation. However, the general results of the model should hold in most play calling situations. Even in most extreme cases, a dominant pure strategy will not occur, and the mixed strategy result should occur. For instance, an offense does not pass 100% of the time when it is third down and 10 or more yards to go for the first down. In these situations the offense passes 85.5% of the time. Likewise, the offense will not run every time they are faced with second down and short yardage. When it was second down and four or fewer yards to go for the first down, the offense ran the ball only 71% of the time, not 100% of the time, as is implied by a dominant pure strategy.[4] The later empirical section will control for different game situations that may influence the payoffs and thus the values of a and y.[5]

Next consider the case when the offense's starting quarterback is injured and is replaced by an inferior passer. The payoffs are summarized in Table 2.

[4] These percentages come from the data set used later in the article.

[5] Let us take another look at whether our results hold in different game situations. We will concentrate on the general result for a, the optimal probability that an offense passes, because the offense is the subject of our later empirical work. Remember $a = (C - D) / [(C - D) + (B - A)]$. Although the payoffs may vary with the game situation (down, yards to a first down, and so on), the basic relationships in the equation above should remain stable. Consider $(C - D)$. In all game situations, C will be greater than D. In this comparison, the offense definitely runs the ball. C is greater than D because the offense will expect to perform better running the ball when the defense is playing against the pass, and comparatively worse when the offense runs the ball against a run defense. A similar story can be told for $(B - A)$. In this case the offense definitely passes the ball. Again, regardless of the game situation, the offense will do better when the pass is thrown against a run defense (B) rather than against a pass defense (A). Although the payoffs may vary with the game situation, C should always remain greater than D, and B should remain greater than A. Therefore, the structural relationship found in "a" should apply in different game situations.

Table 2. Football Play Calling Game with Second String Quarterback

	Defensive Play	
Offensive Play	Defend Pass (y)	Defend Run $(1 - y)$
Pass (a)	$(A - X_1), (-A + X_1)$	$(B - X_2), (-B + X_2)$
Run $(1 - a)$	$C, -C$	$D, -D$

Payoffs: offense, defense.

The second string quarterback will be less proficient at passing than the starter. Therefore, the offense's payoffs will be lower for passing plays compared to the payoffs shown in Table 1 (when the starting quarterback played). Let X_1 be the drop-off in the offensive payoff from passing when the defense is defending the pass. The northwest outcome box shows the new payoff in this situation: A, the payoff obtained with the starting quarterback, minus X_1. Football is a zero sum game, so when the offense does worse, the defense does better. Therefore, we add X_1 to the defense's payoffs for this outcome.

Next consider the case when the defense plays against a run. The substitute quarterback will not be able to exploit this as well as a starting quarterback. The drop-off in the offense's payoffs is given by X_2. X_2 is added to the defense's payoffs in the northeast outcome box.

When the second string quarterback plays, the general mixed strategy equilibrium results change. First, consider the probability that a defense will try to defend against the pass:

$$y = D - (B - X_2)/[\{D - (B - X_2)\} + \{(A - X_1) - C\}]. \tag{3}$$

Equation 3 differs from Equation 1 in a straightforward way. In Equation 3 $(A - X_1)$ replaces A in Equation 1, and additionally $(B - X_2)$ replaces B. Consider the numerator of Equation 3. Even when the substitute quarterback is playing, the offense will still want to pass when the defense is playing against the run, so $D < (B - X_2)$, causing the top part of the fraction to be negative. Next consider the denominator of Equation 3. The first set of terms is identical to the numerator, so it will be negative. Next, consider the second set of terms. When the substitute quarterback is playing, passing becomes even less desirable when the defense plays against the pass, so $(A - X_1) < C$, causing the last set of terms to be negative. Because both components of the denominator are negative, the denominator is negative. The negative signs in the numerator and denominator will cancel, leaving a positive fraction.

Consider the net effect of the substitute quarterback on the probability that the defense will play pass. Because a positive X_2 is added to the negative numerator, and only $(X_2 - X_1)$ is added to the negative denominator, the absolute value of the top of the fraction will decrease by more than the drop in the absolute value of the denominator. Therefore, the probability that the defense plays against the pass, y, must decline. In all, the drop-off in the substitute quarterback's performance will cause the defense to defend against the run more often and defend against the pass less often.

Next, consider the probability that the offense will pass, which is determined by the defense's payoffs. Because the defense's payoffs are negative and it is easy for the reader to lose track of the signs, the calculations for the new probability that the offense will pass, a, appears below.

The payoffs from each action are

$$\text{Payoffs}_{(\text{defend against pass})} = a(-A + X_1) + (1 - a)(-C),$$

$$\text{Payoffs}_{(\text{defend against run})} = a(-B + X_2) + (1 - a)(-D).$$

Table 3. Implications of X_1 and X_2 for Offensive Play Calling

Relationship	Predicted Outcome
$X_1 = X_2$	No change in likelihood of passing
$X_1 > X_2$	Offense passes less often
$X_1 < X_2$	Offense passes more often

X_1 = drop-off in performance when the substitute is playing against a pass defense.
X_2 = drop-off in performance when the substitute is playing against a run defense.

Because the defense is indifferent between the two strategies in equilibrium, we can set the payoffs equal:

$$a(-A+X_1)+(1-a)(-C)=a(-B+X_2)+(1-a)(-D).$$

This can be simplified to

$$-aA+aX_1-C+aC=-aB+aX_2-D+aD,$$
$$-aA+aX_1+aC+aB-aX_2-aD=C-D,$$
$$a(-A+X_1+C+B-X_2-D)=(C-D),$$

so

$$a=C-D/[(C-D)+(B-X_2)-(A-X_1)]. \tag{4}$$

Ignore X_1 and X_2 for a moment. Because $C > D$ and $B > A$, both the top and bottom of the fraction have a positive sign. Notice that only the payoffs in the later part of the denominator are influenced by the substitute's drop-off in performance. Consider the case when the second string quarterback's performance drop-off is the same, regardless of whether the defense plays against the pass or run ($X_1 = X_2$). In this case, X_1 and X_2 will cancel, and the offense will not change its play calling. They will call a passing play the same percentage of the time, regardless of whether its first string or second string quarterback is playing.

If the second string quarterback has the greatest drop-off in performance when the defense is playing against the pass, then $X_1 > X_2$. The numerator becomes larger because the two effects no longer cancel. Consequently, the offense will pass less. In contrast, if a second string quarterback has a larger drop-off when passing against a run defense, $X_2 > X_1$, the numerator becomes smaller, and the offense will pass more often. The relationship between X_1 and X_2 determines the predictions the model makes. The predictions for the offensive play calling are summarized in Table 3.

The remainder of the article empirically examines the game theoretic predictions for offensive play calling. Offensive plays have an advantage over their defensive counterparts, namely, that they are readily observable. It is clear if an offense passed or ran the ball; although, it is much harder to discern if the defense was playing against the run or pass.

3. An Examination of X_1 and X_2

This section attempts to do two things. First, it attempts to empirically examine the assumption that there is a performance drop-off when a substitute quarterback is playing for an

Table 4. Quarterback Performance in Third and Long and Second and Short

	Third and Long	Second and Short
Starter:		
No. first downs	44	26
No. that did not make first down	95	22
Total passes	139	48
% first down obtained	31.7%	54.2%
Average yards per attempt	8.92	5.75
Substitute:		
No. first downs	21	20
No. that did not make first down	82	22
Total passes	103	42
% first downs obtained	20.4%	47.6%
Average yards per attempt	7.25	4.38

injured starting quarterback. Second, it tries to determine whether X_1 is greater than, less than, or equal to X_2.

To accomplish these tasks, we consider the data set used for our regression analysis later in the article. There were 11 teams that had a quarterback who was replaced because of injury in 2006.[6] The data set includes all first half plays for all 16 regular season games for these 11 teams, for a total of 5116 observations.[7] The second half plays were excluded from the data set because the score may dictate whether a team runs or passes in this half. A team that is way ahead may run to take time off the game clock, while a team that is far behind may pass to try and catch up quickly. It is very likely these adjustments are not made in the first half, so by considering only first half plays, we can remove the score of the game as a factor in a team's decision of whether or not to pass.

The payoffs in the model were presented in utility levels. In order to move from the theoretical model to the empirical model, we need measurable proxies for utility. Performance statistics are an obvious candidate for proxies. However, because there is not a single obvious measure of quarterback performance, we take the approach of considering two plausible measures of quarterback performance and of using these to quantify X_1 and X_2. The first performance measure is average yards per passing attempt; the second measure is the percentage of pass attempts that result in first downs.[8]

To quantify X_1 for either performance measure, we need to identify obvious passing downs.[9] We look at all of the third down plays where there were 10 or more yards to go for a first down and where the quarterback passed. Table 4 summarizes the performance measures. In 139 passing attempts, starters averaged 8.92 yards per attempt, and in 103 passing attempts substitute quarterbacks averaged 7.25 yards per attempt. When it was third and long, starters

[6] We found which players were injured by looking at the injury reports on various sports-centered websites. Examples include http://msn.foxsports.com and http://sports.yahoo.com/nfl/news?slug=jm-injuries110906&prov=yhoo&type=lgns.

[7] The play-by-play data and the game summaries were found at Fox Sports (2006) (each game has its own identification number).

[8] Steven Levitt suggested using this second performance measure.

[9] By obvious passing downs we simply mean a situation where a pass is very likely. However, even in these cases, an optimal mixing strategy is likely to suggest that an offense should not pass every time it is confronted with this situation. In these obvious passing downs, offenses passed 85.5% of the time: 87% when starters were playing, and 84% when substitutes were playing.

obtained the first down 31.7% of the time, but substitutes were successful only 20.4% of the time.[10] Both performance measures suggest that there was a drop-off in performance when substitutes were playing in passing downs. However, these differences do not always reach the level of statistical significance. A chi square test of the differences among proportions allows us to reject, at the 7% level, the null hypothesis that starters and substitutes have the same proportion of passes that result in first downs.[11] However, a test of the difference between sample means cannot reject, at conventional levels, that the starters and substitutes have the same average yards per passing attempt.[12]

Next, to determine the size of X_2, we try to identify obvious running situations. We look at all of the pass attempts when it was second down and four or fewer yards to go for a first down. We also exclude plays after the two minute warning because the amount of time left in this case may make even a second and four a passing down. Table 4 also summarizes the two quarterback passing performance statistics in running situations.[13] In 48 passing attempts, the starter averaged 5.75 yards per attempt. In 42 attempts the substitute quarterbacks averaged 4.38 yards per attempt. A similar story is told by the other performance measure. Starters successfully obtained first downs 54.2% of the time, and substitutes met this goal 47.6% of the time. With either performance measure, compared to starters, substitutes are not as proficient in passing in second and short situations. However, in the second and short situations, neither drop-off in performance variables was significant at conventional levels.

Our next task is to determine whether the drop-off in a substitute's performance is greater against the pass or run (is $X_1 >, <,$ or $= X_2$). To test this relationship, we run an ordinary least squares (OLS) regression. The data set includes all third down and long and second down and short passes in the data set (the 332 observations—139 + 48 + 103 + 42—summarized in Table 4).

The equation takes the following form:

$$Y_i = B_1 \times Z_{1i} + B_2 \times Z_{2i} + B_3 \times Z_{3i} + B_4 \times Z_{4i}.$$

The dependent variable Y_i is the yards per attempt on play i, the Bs are estimated coefficients, and the Zs represent the dummy variables. There is one dummy variable for each of the four categories found in Table 4. They are (i) starter on third and long, (ii) starter on second and short, (iii) substitute on third and long, and (iv) substitute on second and short. Z_1 is a 1 for all observations in the first group in the previous sentence, and Z_2 is a 1 for all observations in the second group, and so on. There is no constant, because it would be co-linear with the dummy variables. The estimated regression coefficients will measure each group's performance.[14] Next, consider the difference in some of the

---

[10] In this sample, the average yards to go for a first down were 13.28 for starters and 13.07 for substitutes. Because the two types of quarterbacks found themselves in similar situations, the performance difference is not due to substitutes making riskier passes because they have more yards to go for the first down.

[11] The formula for the difference in proportions when there are only two groups is $\sum[(|f - e| - 1/2)^2/e]$, where f is the observed frequency and e is the expected frequency. See Mansfield (1987, p. 358) for more detail. Our test statistic, which is distributed along a chi square distribution, is 3.27. There is one degree of freedom in this test.

[12] The formula for the difference between sample means is $z = (x_1 - x_2) \Big/ \sqrt{(\sigma_1^2/n_1 + \sigma_2^2/n_2)}$, where x indicates a sample mean, σ^2 is a variance, and n is a sample size. The z-statistic is 1.14.

[13] In obvious running downs, offenses passed 29% of time (29.45% of the time when starters were playing, and 29.37% when substitutes were playing).

[14] Of course, the regression could have been run with a constant if one of the dummy variables was removed. In this case the marginal effect of the dummy variable would be obtained by adding or subtracting the estimated coefficient for each dummy from the estimated coefficient for the constant. This approach is convenient because it allows one to use an F-test to test whether the three remaining dummy variables are jointly significant. When we estimated the model with this specification, we obtained an F-statistic of 2.54, which has a significance level right around 5%.

Table 5. A Test of Whether a Substitute's Drop-off in Passing Performance Is the Same in Passing and Running Situations

	OLS (Yards)	Tobit (Yards)	Probit (Dummy Variable = 1 if First Down Obtained)
F-statistic	0.01 (0.908)		
Chi square statistic		0.17 (0.68)	0.34 (0.56)

The columns indicate the estimation technique and the dependent variable in parentheses. The independent variables are four dummy variables indicating (i) starter on third down and long, (ii) starter on second down and short, (iii) substitute on third down and long, and (iv) substitute on second down and short. The entries in the table test coefficient restrictions. Specifically, they test whether the difference of the coefficients on the first and third dummy equal the difference of the coefficients on the second and fourth dummy variables. The entry represents the test statistic, and the level of significance appears in parentheses.

estimated coefficients. $(B_1 - B_3)$ will capture the performance drop-off when a substitute is passing in an obvious passing situation (third and long). $(B_2 - B_4)$ is the performance drop-off when a substitute is passing in a running situation (second and short). We test the restriction that $(B_1 - B_3) = (B_2 - B_4)$. An *F*-statistic of 0.01 is insignificant at conventional levels. That is, we cannot reject the hypothesis that the drop-off in performance, when substitutes are playing, is the same regardless of whether it is second and short or third and long. This suggests that $X_1 = X_2$.

Table 5 presents the aforementioned *F*-statistic, along with some additional tests of the hypothesis that the substitute quarterback's passing performance drop-off is the same in passing and running situations. The second column deals with a potential problem in the first estimation. The data have some zeros, which represent incomplete passes or completions without a gain. These data could easily be considered censored data, except that several observations are negative because of completions that lost yards. As a way to check the robustness of the previous column's results, we delete the five negative observations and estimate the model using a tobit regression. We again test whether $(B_1 - B_3) = (B_2 - B_4)$. A Wald test suggests that the two types of performance drop-offs are not significantly different at conventional levels (chi squared statistic = 0.17). Finally, the model is altered so that the other performance measure is the dependent variable. The dependent variable is a one if a pass results in a first down and equals zero otherwise. The model is estimated with a probit regression. The Wald statistic to test the coefficient restrictions does not allow us to reject the null hypothesis that the drop-off in performance by a substitute does not vary between third and long and second and short.

These three tests of coefficient restrictions all suggest that $X_1 = X_2$. Because offensive play calling is more easily observed than defensive play calling, the model's predictions for offensive play calling will be the subject of our empirical tests. The testable implication of $X_1 = X_2$ is that a substitute quarterback will be just as likely to pass as a starting quarterback.

This outcome is different than the standard isocost/isoquant analysis that predicts a less productive input will be used less often; in this case an isocost/isoquant analysis suggests that a substitute quarterback is less likely to pass.

4. A Crude Comparison of Backup and Starting Quarterbacks

Table 6 lists the 11 teams whose starting quarterbacks were injured in the 2006 season. The starting quarterback is listed first, the first substitute is listed next, and the third starter

Table 6. A Quarterback Comparison for Teams with an Injured Quarterback in the 2006 Season

Team	Quarterback	% Pass
Browns	Frye	50.6%
	Anderson	70.2%
Buccaneers	Simms	60.2%
	Gradkowski	51.6%
	Rattay	52.0%
Cardinals	Warner	60.3%
	Leinart	54.8%
Chiefs	Green	48.1%
	Huard	47.03%
Dolphins	Culpepper	59.8%
	Harrington	57.3%
	Lemon	43.8%
Eagles	McNabb	64.6%
	Garcia	54.0%
	Feely	59.3%
Jaguars	Leftwich	49.4%
	Garrard	47.8%
Panthers	Delhomme	54.8%
	Weinke	50.6%
Raiders	Brooks	44.9%
	Walter	52.7%
Seahawks	Hasselbeck	57.8%
	Wallace	52.6%
Steelers	Roethlisberger	52.0%
	Batch	50.0%

is listed next, where applicable. The table lists the percentage of first half plays that were passes for each quarterback. The evidence is inconclusive in determining whether starters pass more often. The starter passed more often than the second string quarterback in 9 of the 11 cases. However, the pairwise differences in passing rates between the starter and the second string quarterback are not significant at conventional levels (t-statistic = 0.49).[15] However, the Browns' quarterbacks may be driving the results because the second string quarterback passed so much more often than the starter. When the Browns' quarterbacks are removed from the sample, the pairwise differences show that starting quarterbacks pass more often, and this result is statistically significant at the 5% level in a one-tailed test (t-statistic = 2.10).

It also appears that the quality difference between the first and second string quarterbacks did not influence a team's decision to pass. Consider the difference in quarterback ratings[16] between the starter and second string quarterback and the difference between the passing percentages for the starter and backup. The correlation between these differences is −0.19, suggesting that better starters did not pass more often. This correlation is insignificant at the 10% level in a two-tailed test.

[15] Although the results are not shown, the relationship between the starters and the second string quarterbacks is similar to a comparison of the starter and all of the team's backups considered together.

[16] The quarterback rating can be found at National Football League (2006).

The pairwise comparisons are crude because they do not account for the situation the quarterback finds himself in when the team decides whether to pass or run.[17] The regression analysis that follows looks at the relationship between a team's decision to pass and whether or not the starting quarterback is playing by holding constant the relevant control variables.

5. Econometric Issues

The general form of the equation that we estimate is

$$P(y_{it}=1|x_{it},c_i)=P(x_{it}B+c_i).$$

Consider the subscripts: i indicates the number of groups, or in this case teams, (I = 1 … 11), and t indicates the number of observations per team. $P(y_{it} = 1)$ is the probability that team i will pass the football on play t. X is a vector of all explanatory variables, and c are the unobserved effects in the model that will vary by group. B is a vector of estimated coefficients for the explanatory variables. In the present application, there are two estimation considerations.

The first task is to figure out how to deal with c. One choice is to estimate a fixed effects model that estimates a separate parameter for a series of dummy variables that indicate each group. Because every group has a dummy variable, the equation does not have a constant. The second approach is a random effects model that moves c to the error term. The present model fits the fixed effects specification better than a random effects specification. A fixed effects model may have two problems that are addressed by using a random effects specification. The first problem is that the fixed effects model assumes away the effects of any variables that do not vary with time. Common variables that fit this description are race and religion.[18] However, because all of the explanatory variables are time dependent (yards to go for first down, etc.), this potential problem does not apply to this model. The second problem that may occur in a fixed effects model is that a separate coefficient is estimated for each group (team), resulting in a loss of degrees of freedom. However, the problem is not severe in this case because there are only 11 teams in the sample and over 400 plays for each team. The final estimates are done with over 5000 degrees of freedom.

The second task is to decide which estimation technique to use. The dependent variable in the model is a binary response (1 or 0), which is often estimated with a probit or logit model. In both the fixed effects probit model and the fixed effects logit model, estimating B and c_i create an incidental parameters problem that leads to an inconsistent estimate of B. This is a small sample bias that exists because the asymptotic properties of the estimator do not converge to the true value of B as the number of groups increases because the number of observations per group is considered to be fixed. When the observations per group are very small, the bias can be very large. When there are two observations per group, Greene (2004b) reports that the bias

[17] Consider two other crude comparisons. In the first half plays mentioned earlier in this section, starters passed 54.4% of the time, and substitutes passed 52.9% of the time. These percentages are derived from the 2755 plays run by starting quarterbacks and 2361 plays run by starters. Chi square tests cannot reject the null hypothesis that substitutes and starters pass the same proportion of the time. Also, on first down plays, a chi square test suggests the proportion of passes by starters and substitutes are not significantly different at the 0.05 level.

[18] See Kennedy (2003).

reaches 100%. However, Greene (2004b) finds that the bias shrinks as the number of observations in a group increases. Greene performs Monte Carlo estimates to compared random effects, fixed effects, and pooled estimators of a probit model. He concludes that the fixed effects probit estimator is the best choice once the number of observations in each group gets above eight.[19] Our sample has over 400 observations per team, which is a big enough sample that we can ignore the small sample properties of the estimators.

One variation of the logit model has been developed that eliminates the fixed effects bias. The conditional logit is a technique that is similar to the linear estimate of fixed effects in the following sense. When OLS is used to estimate a fixed effects model, a computational trick is used to make the task of estimation easier. The data are transformed by subtracting the average of each variable within a group from the individual observation. Therefore, each fixed effect parameter does not have to be estimated by a computer program. The same approach allows the conditional logit to construct an estimate of B that is independent of the fixed effects, c.[20] However, several drawbacks preclude us from using the conditional logit. Most importantly, estimation is not yet feasible. LIMDEP claims that the program can estimate the conditional logit only when the number of observations is no more than 20 or 30 observations per group. In any event, the absolute limit on the size of the groups is 100, and at this size the program will probably not be able to perform the estimation.[21] Because our sample has over 400 observations per group, this technique cannot perform the estimation task. Katz (2001) and Coupé (2005) conclude that when the number of observations per group exceeds 16, the bias of an unconditional logit (which has dummy variables) is small. An additional disadvantage of the conditional logit is that it does not estimate the fixed effects parameters, so the marginal effects cannot be examined.[22]

Given these considerations, we present estimates of a fixed effects probit as well as estimates of a pooled probit. A pooled probit is similar to a fixed effects probit with the exception that the dummy variables measuring the fixed effects are not included in the model. Presenting both estimates has the following advantage. In Monte Carlo experiments, Greene (2004b, p. 110) finds that the fixed effects model will estimate the coefficients with an upward bias; whereas, the pooled probit will estimate the coefficients with a downward bias. The true values of the parameters should be somewhere in between the two types of estimates that we report. Fixed effects logit and pooled logit are also a plausible estimation strategy. We also estimate the empirical model with these logit techniques and find these estimates are very similar to the estimates produced by their probit counterparts, so only the probit estimates are reported.

6. Estimated Equation and Results

Definitions of the variables in the equation to be estimated are the following: $Pass_{it} = f(2nd\ Down_{it},\ 3rd\ Down_{it},\ ToGO1st_{it},\ ToGOTD_{it},\ Two\ Minute\ Warning_{it},\ SQ_{it})$. *Pass* is a dummy variable that takes the value of 1 if team i passes the ball on play number t and is a zero otherwise. *2nd Down* and *3rd Down* are dummy variables if it is a second down or a third down play, respectively. *ToGo1st* is the number of yards the offense must go to get a first down.

[19] See Greene (2004b, p. 111).
[20] See Wooldridge (2002, p. 500–504) for a discussion.
[21] See Greene (2002), E16–12.
[22] See Greene (2002), E16–14.

Table 7. Probit Estimates of the Offense's Decision to Pass

	1	2	3
Constant	−0.910 (−12.97)		
2nd Down	0.308 (7.28)	0.312 (7.36)	0.367 (5.81)
3rd Down	1.107 (20.43)	1.117 (20.53)	1.159 (14.72)
ToGO1st	0.059 (10.71)	0.060 (10.80)	0.061 (7.53)
TOGOTD	0.0013 (1.63)	0.0014 (1.69)	0.0008 (0.71)
Two Minute	0.645 (11.20)	0.642 (11.11)	0.620 (7.28)
Starting quarterback (SQ)	0.052 (1.42)	0.044 (1.04)	0.055 (0.40)
SQ*2nd Down			−0.100 (−1.17)
SQ*3rd Down			−0.078 (−0.71)
SQ*TOGO1st			−0.002 (−0.20)
SQ*TOGOTD			0.001 (0.62)
SQ*Two Minute			0.042 (0.36)
Log likelihood function	−3229.26	−3212.02	−3211.01

Column 1 presents pooled probit estimates. Columns 2 and 3 present fixed effects probit estimates.

TOGOTD is the number of yards the offense has to go to get a touchdown. *Two Minute Warning* is a dummy variable equal to one if the play occurred after the two minute warning. *SQ* is a dummy variable if the starting quarterback ran the play.

As a starting point consider column 1 of Table 7, which presents the pooled probit estimate of the model. In this estimate no consideration is given to any adjustment to accommodate the panel structure of the data. The control variables all move in the expected direction, and all but yards to go for a touchdown are significant at conventional levels. Column 2 presents the fixed effects probit estimate of the same model. The results are qualitatively very similar to the first estimation. Most importantly, both estimates indicate that a starting quarterback is not significantly more likely to pass than his backup. Keep in mind that the fixed effects probit has a bias that overestimates the coefficient and underestimates the standard error.[23] Both biases will serve to inappropriately inflate the *t*-statistics. Even with the inflated *t*-statistic, the *t*-ratio for the quarterback dummy is only 1.04.

The final column in Table 7 adds several interaction terms. The starting quarterback dummy variable is multiplied by each variable in the model. These interaction terms are added to the fixed effects model. None are significant at even the 20% level. A likelihood ratio test also suggests that these interaction terms are jointly insignificant at conventional levels. This result suggests that the estimated coefficients for the model do not vary when a starting quarterback is playing rather than a backup.[24,25]

[23] See Greene (2004a, pp. 136–7; 2004b).

[24] We ran several tests to check the robustness of our results. We found that even with different specifications of the model, changing the starting quarterback did not change the likelihood that a team would pass. We ran a Likelihood Ratio (LR) Test to test the joint significance of the interaction terms. The LR statistic of 2.02 is less than the critical value at the 5% level of 11.07. We also estimated the model by including only one interaction term in the model at a time. In all five cases, the quarterback dummy variable remained insignificant at the 5% level. The individual interaction terms never had a *t*-statistic greater than the absolute value of 1.01. We also divided the data set by downs. Next, we estimated the fixed effects model (without interaction terms) using data from first down plays; then we estimated the model on second down plays; and, finally, we estimated the model on third down plays. In all three cases the absolute values of the *t*-statistics were less than 0.94.

[25] An anonymous referee has noted that it would be interesting to analyze how blitzing influences our results. Unfortunately, our data come from print sources that do not indicate defensive plays. The influence of blitzing on the model is left for future research.

Table 8. Probit Estimated Coefficients and *t*-Statistics When the Model Is Estimated Separately for Each Team

Team	Coefficient	*t*-Statistic
Browns	−0.531	−3.17
Buccaneers	0.274	1.69
Cardinals	0.076	0.56
Chiefs	0.101	0.78
Dolphins	−0.025	−0.17
Eagles	0.307	2.51
Jaguars	0.070	0.60
Panthers	0.121	0.76
Raiders	−0.154	−1.21
Seahawks	0.068	0.49
Steelers	0.061	0.25

This table presents the estimates of the model reported in column 1 of Table 7. The model is run for the data of each of the 11 NFL teams separately.

Next we split up the panel data set and estimate the model separately for each team. In all, there are 11 probit estimates of the model—one for each team. To save space, only the estimated coefficients and *t*-statistics for the quarterback dummy variables are listed in Table 8. The estimates of the separate regressions tell the same story that we found in the full panel data set. NFL teams are not more likely to pass when they have a starter in the game. In only 2 of the 11 cases does a starting quarterback have a greater likelihood of passing (than his substitute) that is significantly different from zero at the 5% level in a one-tailed test.[26]

Considering all the empirical work together, it appears that football teams are very sophisticated. They are calling offensive plays as if they were adopting a mixed strategy game theoretic approach. It should come as no surprise that NFL decision making is so sophisticated because it is a billion dollar business. With so much money at stake, teams have every incentive to call plays in the way that gives them the best chance for success.

7. Test of an Alternate Hypothesis

Our results suggest that an offense will not change the frequency with which it calls pass plays when the starting quarterback is injured. The offense will not change its play calling because it

[26] We check the robustness of the model's estimated results in several ways. First, we run separate estimates of the model for each team: (i) without interaction terms and then (ii) with the interaction terms that appear in Table 7. For the two estimates for each team, we then calculate a likelihood ratio statistics to test the joint significance of the interaction terms. The LR statistics are insignificant at the 5% level in 9 out of the 11 cases. Only the regressions for the Chiefs and the Dolphins have jointly significant interaction terms. We also estimate the model for each of the 11 teams when we include only one interaction term at a time. We do this for all five interaction terms. Only 3 of the 55 estimated interaction term coefficients are significant at the 10% level. These results suggest that the coefficient estimates are stable across the different quarterbacks. Next, we remove the dummy variables for second and third downs and estimate the model (without interaction terms) separately using data exclusively from (i) the first down plays, (ii) the second down plays, and (iii) the third down plays. Among the 11 teams there are 33 estimates for the starting quarterback dummy. Two estimates are positive and significant with *t*-statistics greater than 1.65. Three estimates are negative and significant with *t*-statistics less than −1.65. The remaining 28 estimates are insignificant at conventional levels. These results confirm our result that starting quarterbacks were not more likely to pass on any particular down.

expects the defense to play against the run more often. We consider another hypothesis that may explain why play calling does not change in response to a quarterback injury. Namely, a coach's system may be more important for play calling than which players are available to the coach. An offensive system may be too complicated to respond to a random injury shock. We can analyze this question by identifying a case when a head coach leaves midseason. A new coach may call a different mix of pass/run plays if he wishes, but he will still inherit the former coach's system. If play calling changes with the new coach, then the system is not so complicated that it prevents possible changes in play calling. No coach was replaced during the 2006 season analyzed in this article. However, Bobby Petrino resigned as the Atlanta Falcons head coach after 13 games in the 2007 season. The final three games were coached by Emmitt Thomas.

During the 2007 season, the Falcons played three quarterbacks. Joey Harrington started the season and played games 1–6, 8–9, and 11–12. Byron Leftwich, who was acquired after the second game of the season, played in the seventh and tenth games. Chris Redman played one play in the ninth game because Harrington was out with an injury for one play. Redman also played in the last four games. The quarterback changes had more to do with the starters' ineffectiveness than it did with injuries.[27]

We ran our model on all of the first half plays for the Falcons 2007 season. We report the results from various specifications of our model in Table 9. The results on this issue are mixed. First, consider the evidence that suggests that a coaching change did not alter play calling. In column 1 the model includes a dummy variable for the games Petrino coached and no dummy variables for the quarterbacks. The coach dummy variable is insignificant at conventional levels. Next, we split the sample into games coached by Petrino and those he did not coach. We run separate estimates of the model in column 1 (without the coaching dummy variable). A likelihood ratio statistic is 5.84, which is less than the critical value of 11.07 at the 0.05 level. Therefore, the likelihood ratio test cannot reject the null hypothesis that the model's coefficients are equal across the two samples.

Next, consider the evidence that suggests that a coaching change influenced play calling. Column 2 contains a dummy variable for the games coached by Petrino, as well as a dummy variable for the games started by Redman, who is the only quarterback to start games for both Petrino and the interim coach. The third column estimates the model with a coaching dummy variable and a dummy variable for the season starter Harrington. Finally, in column 4, the model is estimated with a coaching dummy variable and a dummy variable for Harrington, as well as one for Redman. The coefficient on the Petrino dummy variable is positive in all three cases, suggesting that the Falcons were more likely to call a passing play when they were coached by Petrino. The t-statistic on the coaching dummy variable ranges from 1.56 to 1.97. The levels of significance range from 5% to 12%. This suggests that when quarterback play is held constant, an offense can alter its play calling. Therefore, the existing system can be changed if the coach wishes to change it.

8. Serial Independence

In mixed strategy equilibrium games, the players try to be unpredictable. To be unpredictable, an offense must randomly choose between pass and run. That is, the play call

[27] For a description of the game and a discussion of why quarterbacks were changed during the season see http://www.nfl.com/gamecenter/recap?game_id=28875&displaypage=tab_recap&season=2006&week=reg1.

Table 9. Probit Estimates of the Atlanta Falcons Offense's Decision to Pass in the 2007 Season

	1	2	3	4
Constant	−1.170 (−4.29)	−1.507 (−4.01)	−1.171 (−4.28)	−1.258 (−3.13)
2nd Down	0.09 (0.68)	0.0960 (0.690)	0.090 (0.65)	0.091 (0.65)
3rd Down	1.23 (6.90)	1.242 (6.92)	1.251 (6.94)	1.252 (6.94)
ToGO1st	0.082 (4.46)	0.082 (4.45)	0.081 (4.40)	0.081 (4.40)
TOGOTD	0.0027 (1.03)	0.0025 (0.94)	0.0028 (1.06)	0.0027 (1.03)
Two Minute	0.247 (1.20)	0.217 (1.05)	0.259 (1.25)	0.251 (1.20)
Petrino dummy	0.145 (0.88)	0.469 (1.58)	0.401 (1.97)	0.465 (1.56)
Redman dummy		0.349 (1.31)		0.090 (0.30)
Harrington dummy			−0.338 (−2.17)	−0.312 (−1.75)
Log likelihood function	−295.18	−294.31	−292.80	−292.76

The last three variables are dummy variables that equal one when (i) Bobby Petrino was head coach, (ii) Chris Redman was quarterback, and (iii) Joey Harrington was quarterback.

must be a random draw, based on the optimal probability the offense chooses for passing. If the defense sees a pattern, it will exploit the pattern to the detriment of the offense. This section tests whether play calling was random as mixed strategy equilibrium play requires.

First, in order to analyze play calling, we need to identify comparable situations faced by the offense. We choose the play calls when the offense faces a first down and 10 yards to go for another first down. We use all of these plays from the first half of the games in the 2006 season for the 11 teams considered in this article. Next, we perform a Runs Test on the play calling. A Run is a string of consecutive play calls that are the same. If there are too many Runs, the last offensive play call is negatively correlated with the previous play call. If there are too few Runs, the last offensive play call is positively correlated with the previous play call. The defense can use the positive or negative correlation in offensive play calling to make adjustments to more effectively foil the offensive plays. If the offense calls plays randomly, it does not give the defense information unnecessarily, allowing the offense to retain its unpredictability.

The results from the Runs Test appear in Table 10. We are able to reject the null hypothesis of random play calling for the Arizona Cardinals, but not for the remaining 10 teams in our data set. This result fits into the sparse literature on tests of mixed strategy equilibriums tested outside of a laboratory. The evidence is mixed concerning whether tennis players choose the side to serve on randomly. Walker and Wooders (2001) find too many Runs in their data, suggesting that players alternate serves too often. They note that "There is overwhelming empirical evidence that when people try to generate 'random' sequences they generally switch too often to be consistent with randomly generated choices" (2001, p. 1533).[28] Using a different data set, Hsu, Huang, and Tang (2007) report that tennis serve location appears to be random. In soccer penalty kicks, Chiappori, Levitt, and Groseclose (2002) and Palacios-Huerta (2003) suggest that the direction of penalty kicks and goalie dives to block the kick are serially independent.

These results, along with our own, tell an interesting story. In tennis a player must both perform on the field and determine strategy. Further, the many decisions on serve location occur during one match. In soccer, the player has to perform on the field and probably has to

[28] See Wagenaar (1972) for a review of the psychology literature on this subject. Walker and Wooders (2001) cite this article.

Table 10. Test for Serial Independence of Play Calling for First Down and 10 Yards to Go for the First Down

				Level of Significance	
Team	Total Plays	Passing Plays	No. of Runs in Data	*Z*-Statistic	Two-Tailed
Browns	178	80	92	0.44	0.658
Buccaneers	176	67	77	−1.12	0.262
Cardinals	181	89	108	2.46	0.014**
Chiefs	172	65	78	−0.63	0.529
Dolphins	192	87	96	−0.02	0.982
Eagles	195	113	97	0.14	0.887
Jaguars	189	76	99	1.08	0.280
Panthers	190	95	99	0.436	0.663
Raiders	174	60	84	0.74	0.461
Seahawks	205	105	105	0.22	0.827
Steelers	195	85	100	0.45	0.651

This table reports the results of runs test performed on first down plays with 10 yards to go for the first down. The data include all first half plays from all 16 games for each team listed above. ** indicates the test statistic is significant at the 0.05 level. The null hypothesis is randomness in play calling.

make his own decision about which side to kick the ball. However, the player has more time to think about the decision because penalty kicks may happen in different games, played on different days.[29] In football, the player is responsible for the performance on the field, and the offensive coordinator or head coach is responsible for determining strategy. The separating of the playing duties from the strategic decisions probably increases the likelihood of effective strategic play in football.

9. Conclusion

Professional sports offer a convenient laboratory for testing whether people adopt mixed strategies. It has several advantages over a typical laboratory setting. First, the players and coaches know the game well and have usually spent several years learning the game. Second, players and coaches are highly motivated. Their livelihood depends upon their performance, and they will seek to adopt the best strategies available to them.

We find that NFL offenses call plays as if they were adopting a mixed strategy. The mixed strategy game presented in this article predicts that when there is an injury to a starting quarterback, the offense will not change its play calling, but instead the defense will play against the run more often. Using NFL data from the 2006 season, our evidence suggests that offenses did not change the likelihood that they would pass when the backup quarterback was pressed into duty because of an injury to the starter. This result differs from the predictions of standard optimizing techniques that ignore the actions of rivals. For instance, isocost and isoquant analysis would predict that a coach would call fewer passes when his starting quarterback was injured and had to be replaced by an inferior back-up.

[29] Chiappori, Levitt, and Groseclose (2002) make this point in footnote 20 on page 1147.

This article also finds that play calling is serially independent. This fits into the existing literature in an interesting way. Walker and Wooders (2001) suggest that tennis players do not randomize their serve placement. These inefficient strategic choices may occur because serve choices must be made by the player in rapid succession as a game unfolds. In football the playing duties are separated from the strategic decision making. A specialist, the coach, calls the plays. It is not surprising that the coach can randomize better than a tennis player, who must perform both the playing and strategic roles.

The state of the literature is in an interesting place with respect to how applicable mixed strategies are to actual decision making. The results are in from one extreme case and are starting to emerge from the other extreme case. In the laboratory setting when players do not know the game very well and they had little at stake, mixed strategies do not seem to be employed. In professional sports, where participants have enormous amounts of money at stake and excellent training, this article, along with Chiappori, Levitt, and Groseclose (2002) and Walker and Wooders (2001), suggests that people do employ mixed strategies. Two tasks remain for future research. First, because our empirical work is one of very few tests of the mixed strategy equilibria done outside of the laboratory, the second extreme case mentioned above needs more empirical scrutiny to make sure that the results hold. Second, the in-between cases need to be examined outside of the laboratory.[30] That is, will the mixed strategies hold when the players have less money at stake than professional athletes and coaches have at stake? Will the results hold when people do not know the game as well as professional athletes and coaches? It is an interesting question whether our results would hold up in college football where less money is at stake, or in high school where there may not be any money at stake.

References

Bradbury, John Charles, and Douglas Drinen. 2007. Crime and punishment in Major League Baseball: The case of the designated hitter and hit batters. *Economic Inquiry* 45:131–44.

Camerer, Colin F. 2003. *Behavioral game theory: Experiments in strategic interaction*. Princeton: Princeton University Press.

Chiappori, Pierre-Andre, Steven Levitt, and Timothy Groseclose. 2002. Testing mixed strategy equilibria when players are heterogeneous: The case of penalty kicks in soccer. *American Economic Review* 92:1138–51.

Coloma, German. 2007. Penalty kicks in soccer: An alternative methodology for testing mixed strategy equilibria. *Journal of Sports Economics* 8:530–45.

Coupé, Tom. 2005. Bias in conditional and unconditional fixed effects logit estimation: A correction. *Political Analysis* 13:292–5.

Fox Sports. 2006. NFL GameTrax. Accessed January 30, 2007. Available at http://msn.foxsports.com/nfl/playbyplay?gameId=20060910005 (each game has its own identification number).

Game Center. 2006. Accessed September 2, 2008. Available at http://www.nfl.com/gamecenter/recap?game_id=28875&displaypage=tab_recap&season=2006&week=reg1 (each game has its own identification number).

Goff, Brian, Robert McCormick, and Robert Tollison. 2002. Racial integration as an innovation: Empirical evidence from sports leagues. *American Economic Review* 92:16–26.

Goff, Brian, William Shughart, and Robert Tollison. 1997. Batter up! Moral hazard and the effects of the designated hitter rule on hits batsman. *Economic Inquiry* 35:555–61.

Greene, William. 2002. *LIMDEP econometric modeling guide: Version 8.0*. Plainville, NY: Econometric Software.

Greene, William. 2004a. Fixed effects and bias due to the incidental parameters problem in the Tobit model. *Econometric Reviews* 23:125–47.

Greene, William. 2004b. The behavior of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *Econometrics Journal* 7:98–119.

[30] Walker and Wooders (2001) also make this point.

Heckelman, Jac, and Andrew Yates. 2003. And a hockey game broke out: Crime and punishment in the NHL. *Economic Inquiry* 41:705–12.

Hsu, Shih-Hsun, Chen-Yin Huang, and Cheng-Tao Tang. 2007. Minimax play at Wimbledon: Comment. *American Economic Review* 97:517–23.

Katz, Ethan. 2001. Bias in conditional and unconditional fixed effects logit estimation. *Political Analysis* 9:379–84.

Kennedy, Peter. 2003. *A guide to econometrics.* 5th edition. Cambridge, MA: MIT Press.

Klaassen, Franc, and Jan Magnus. 2009. The efficiency of top agents: An analysis through service strategy in tennis. *Journal of Econometrics* 148:72–85.

Mansfield, Edwin. 1987. *Statistics for business and economics*. New York: W. W. Norton and Company.

McCormick, Robert, and Robert Tollison. 1984. Crime on the court. *Journal of Political Economy* 92:223–35.

National Football League. 2006. NFL.com. Accessed January 30, 2007. Available at http://www.nfl.com/stats/player?seasonid=2006&seasontype=reg&submit=go.

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

Poitras, Marc, and Lawrence Hadley. 2006. Do new major league ballparks pay for themselves? *Journal of Business* 79:2275–99.

Sutter, Daniel, and Stephen Winkler. 2003. NCAA scholarship limits and competitive balance in college football. *Journal of Sports Economics* 4:3–18.

Wagenaar, W. A. 1972. Generation of random sequences by human subjects: A critical survey of literature. *Psychological Bulletin* 77:65–72.

Walker, Mark, and John Wooders. 2001. Minimax play at Wimbledon. *American Economic Review* 91:1521–38.

Wooldridge, Jeffrey M. 2002. *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.