Are NFL Coaches Risk and Loss Averse? Evidence from Their Use of Kickoff Strategies

John D. Urschel, State University of New York at Buffalo
Jun Zhuang, State University of New York at Buffalo

Recommended Citation:
Available at: http://www.bepress.com/jqas/vol7/iss3/14
DOI: 10.2202/1559-0410.1311
©2011 American Statistical Association. All rights reserved.
Are NFL Coaches Risk and Loss Averse?
Evidence from Their Use of Kickoff Strategies

John D. Urschel and Jun Zhuang

Abstract

Quantitative analysis of football play calling suggests that NFL coaches do not choose their strategies optimally. They tend to be overly cautious. One possible explanation for this finding is that NFL coaches are averse to risk and loss. We propose a prospect theory based model of coaches' utility and estimate the model's parameters using kickoff data from the 2009 NFL season. Using an outcome measure of points scored on the initial post-kickoff possession we analyze two strategic kickoff decisions that involve risk-reward tradeoffs: the decision to kick a surprise onside kickoff or a regular kickoff, and the decision to accept a touchback or run the ball out of the endzone. Surprise onside kickoffs may yield a more favorable mean points scored value for the kicking team than a regular kickoff, yet surprise onside kickoffs are infrequently used (and thus the same size is small and the p-value of significance test is 0.68). Coaches appear averse to the possible loss involved in the surprise onside kickoff. Running the ball out yields a higher mean points scored for the receiving team than accepting a touchback, but it entails some risk (fumbles are lost in 2 percent of returns). Nevertheless, declining the touchback option and running the ball out is very common. Coaches do not appear excessively risk averse when presented with this choice over possible gains. Prospect theory models allow for risk aversion over possible gains, as in traditional expected utility theory, and in addition they permit an asymmetric aversion to losses. A prospect theory model therefore seems suitable for our analysis of kickoff strategies. We estimate a risk aversion coefficient value of 0.66 and a loss aversion coefficient value of 1.55, where values <1 and >1 indicate risk and loss aversion, respectively. Our analysis supports the notion that NFL coaches are both modestly risk averse and loss averse. In other words, coaches display diminishing sensitivity to changes in scoring outcomes as they move further from a reference point (zero), and for scoring gains and losses of equal magnitude they suffer more from a loss than they enjoy from a gain. This result may explain their propensity for making conservative strategic choices that, at first glance, appear sub-optimal.

KEYWORDS: risk aversion, loss aversion, utility, prospect theory, football
1. Introduction

Quantitative analysis of the strategic decision making of National Football League coaches has generally yielded unflattering conclusions. Coaches appear to underutilize the pass despite higher expected rewards compared to running (Alamar, 2006; Alamar, 2010; Kovash and Levitt, 2009), and they appear overly cautious in 4th and short situations (Romer, 2006). This seemingly conservative and suboptimal play is puzzling given the competitiveness of the NFL and the large rewards for winning. One possible explanation for this apparent puzzle is that coaches do not have neutral preferences with respect to risk and loss; their utility functions over uncertain outcomes may be non-linear. In other words, coaches may actually be making optimal strategic decisions in the context of some unknown set of preferences over risk and loss.

We explore the preferences of coaches over risk and loss using the framework of prospect theory. Roughly speaking, prospect theory is an extension of traditional expected utility theory (von Neumann and Morgenstern, 1944; Kahneman and Tversky, 1979; Tversky and Kahneman, 1992). In both theories a utility function is posited that represents an individual's preference over a range of uncertain outcomes. In an expected utility model individuals are typically assumed to be risk averse over possible gains. That is, the utility function is concave over gains (in expected utility theory risk aversion can be defined by this concavity of the utility function). In effect, the individual displays diminishing utility gains as he moves further away from a reference point (zero in this presentation). A coach, for example, may gain greater utility from a play that results in a 7 point lead compared to a play that increases a lead from 21 to 28 points. Prospect theory considers utility (and disutility) over both gains and losses. Simply extending expected utility concepts to the negative outcome domain gives a symmetrical S-shaped utility function (in the typical risk averse case). Prospect theory models go beyond this. These models allow individuals to be loss averse; they are more sensitive to losses than to gains. For a coach the disutility of a play that gives the opponent a touchdown may exceed the utility of a play that gives his team a touchdown. His S-shaped utility function is asymmetric. It is steeper over the negative domain.

In this paper we use data on kickoff strategies to make inferences about the preferences of coaches over risk and loss. For tractability we treat NFL coaches as a homogeneous group that shares the same preferences. Data from the 2009 NFL regular season and playoffs is analyzed. Two kickoff decisions that involve risk-reward tradeoffs are considered; one for the kicking team and one for the receiving team. First we examine the choice between kicking a surprise onside kickoff or a regular kickoff. Then we examine the choice between accepting a touchback when given the opportunity, or running the ball out of the
endzone. We posit a simplified two parameter prospect theory model of coaches' preferences and estimate the parameters using the kickoff data.

The paper is organized as follows. We start with an empirical analysis of kickoff data. In the methods section we describe the data collection process, and define terms and conventions that are used throughout the paper. We summarize kickoff data and report salient findings in the results section. We present histograms of kickoff outcomes in terms of field position after kickoff, and points scored on the initial post-kickoff possession. In the modeling section we introduce a prospect theory model of coaches' preferences over uncertain outcomes and estimate model parameters. We summarize the paper's findings in the final section.

2. Methods and Definitions

We watched all regular and post-season games from the 2009 NFL season using the NFL's Game Rewind (https://gamewind.nfl.com/nflgr/secure/registerform). Kickoffs were reviewed and data was collected on the type of kick, game situation, and the resulting outcomes. Major outcomes of interest were possession of ball, availability of a touchback option, starting field position, and points scored on the initial post-kickoff possession. In the definitions that follow we refer to the two teams involved in a kickoff as the kicking and receiving teams.

The kicking team is said to use a deep kickoff if the ball is kicked as deep as possible. An onside kick (or kickoff, we use the two terms interchangeably) is a kick designed to travel 10-15 yards with the expressed aim of the kicking team gaining possession. The ball is typically driven into the ground as it comes off the kicking tee as this eliminates the option of a fair catch by the receiving team. A squib kick is defined as a low trajectory, often bouncing on the ground, type kick of intermediate distance. A popup kick is a high in the air, prolonged hang time, type kick of intermediate distance. The squib and popup kicks are often employed to limit the possibility of a long return; the kicking team would prefer that they are fielded by someone other than the two kick return specialists. It is often difficult to distinguish a poorly executed deep kick from a squib or popup kick. We can simplify the kickoff classification into just two groups: onside and regular, where the latter includes deep, squib and popup kicks.

The receiving team is said to have declined a touchback option if it fields a kickoff in the end zone (we define end zone to include the goal line itself), and then elects to run the ball out. A touchback occurs if the receiving team accepts a touchback (the converse of the previous definition) or the ball cannot be fielded within the end zone. It is often difficult to distinguish these two touchback cases. For example, a team may implicitly accept the touchback option by not
attempting to field the ball as it travels through the end zone. In this situation it is difficult for an observer to separate unwillingness from inability to field the ball.

*Field position* is measured on a 0-100 yard integer scale, with zero corresponding to the goal line of the team with ball possession. We round the ball's position to the nearest integer yard line. Returning a kickoff for a touchdown corresponds to a field position of 100 yards. The set of all possible field positions consists of two subsets of integers \( \{0, ... , 100\} \), one for the receiving team (the usual case) and one for the kicking team (if kicking team recovers an onside kick or a fumbled return). This subdivision of field position into two subsets poses a problem for the use of field position as a kickoff outcome. It prevents us from calculating all inclusive summary statistics (such as estimates of central tendency) for field position. Our next outcome measure, points scored on the initial post-kickoff possession, does not exhibit this limitation.

The *initial post-kickoff possession* is defined to include all plays up until the point of scoring (and conversion), or the opposing team starting with a first-down possession. We are interested in the *points scored on the initial post-kickoff possession*, where positive points reflect points scored by the receiving team and negative points indicate points scored by the kicking team. To illustrate the definition of the initial post-kickoff possession and show how points are assigned we consider several examples. If the receiving team scores a touchdown on its initial post-kickoff possession it is credited with either 6, 7, or 8 points, depending on the outcome of the point after touchdown or two-point convert. If the receiving team loses a fumble or is intercepted, and the defensive team scores a touchdown on that same play, an outcome of negative 7 (or -6 or -8) points is assigned. Similarly, if the receiving team punts on the initial post-kickoff possession and the punt is returned for a touchdown, negative 7 (or -6 or -8) points are recorded. If the receiving team gives up a safety on the initial post-kickoff possession negative 2 points is recorded. A successful onside kickoff that leads to a field goal for the kicking team gives an outcome of negative 3 points. It should be apparent that the range of all possible points scored on the initial post-kickoff possession is the set \( \{-8, -7, -6, -3, -2, 0, 2, 3, 6, 7, 8\} \).

Onside kickoffs are sub-classified based on the game setting. The onside kick is deemed a *surprise* if: (i) an experienced football observer, such as a television analyst, finds its use surprising, (ii) the game score and time remaining do not suggest its use, and (iii) the receiving team does not employ its specialized onside (hands) team. All other onside kickoffs are classified as expected. In our review of kickoffs we had remarkably little difficulty distinguishing between these two onside kickoff subtypes. The prototypical example of a surprise onside kickoff (probably for many years to come) is the onside kickoff successfully executed by New Orleans to start the second half of Super Bowl XLIV. To say
that Indianapolis was surprised by that kickoff is an understatement; the front line players were running back towards their blocking positions even before the ball left the kicking tee.

3. Results

3a. General kickoff results

During the 2009 NFL regular and playoff seasons there were 267 games and 2592 kickoffs. Of these kickoffs, 2523 were kicked from the 30 yard line and 69 were kicked from other positions (as the result of penalties on the previous scoring plays or safeties). Penalties occurred on 163 kickoffs. Eighteen kickoffs were returned for touchdowns.

Ninety-eight percent (2543) of kickoffs were regular (deep, squib, and popup) and two percent (49) were onside. Of the 49 onside kickoffs, 13 were surprise kicks and the other 36 were expected. One percent of regular kickoffs (26 of 2543) were fumbled and recovered by the kicking team. Twenty-five percent of onside kickoffs (12 of 49) were successfully recovered by the kicking team. The success rate of surprise onside kickoffs was higher than for expected onside kickoffs (5 of 13 versus 7 of 36), but this difference was not statistically significant (p=0.259, Fisher's exact test). Seventeen percent (422) of regular kickoffs resulted in touchbacks. In an additional twenty-four percent (615) of regular kickoffs the receiving team declined the option of taking a touchback, and ran the ball out of their end zone. The receiving team lost a fumble in two percent of these cases (12 of 615).

3b. Points scored and field position

In this paper we use points scored on the initial post-kickoff possession as our major outcome of interest. Field position is a secondary outcome of interest for us, but it has a prominent role in the related literature (Carter and Machol, 1971; Caroll et al., 1988). In addition, field position is a very natural and intuitive outcome measure for most football analysts. Therefore, we report field position outcomes in some detail here and provide a related measure, expected points for a given field position.

In particular, we estimated field position and points scored on the initial post-kickoff possession independently. That is, points scored was not derived from field position. Other researchers have focused on field position as the major outcome, and then mapped field position to expected points using a conversion function. The conversion factor is often based on the field position to points mapping published by Carter and Machol (1971), which was based on data from
the 1969 NFL season. To assist researchers who prefer this field position to points mapping we grouped the post-kickoff field positions into ten bins (0-9 yards, 10-19 yards, ... 60-69 yards, 70-79 yards, 80-99 yards, 100 yards), and then estimated the expected points scored on the initial post-kickoff possession for each bin. We combined the 80-99 yard interval into one bin because of the low counts for these field positions. A field position of 100 yards, meaning a touchdown scored on the kickoff, was assigned its own bin since its expected point value was essentially deterministic (assuming the usual point after attempt outcome). The results are presented in Table 1. These results for kickoff data from the 2009 NFL season are broadly consistent with more general field position to expected points mappings published by others (Alamar, 2010; Caroll et al., 1988; Kovash and Levitt, 2009).

Table 1. Mapping field position of the receiving team (after a regular kickoff) to expected points scored during the initial post-kickoff possession.

<table>
<thead>
<tr>
<th>Field Position Bin (yards)</th>
<th>Number of cases</th>
<th>Mean Points Scored</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-9</td>
<td>32</td>
<td>0.88</td>
</tr>
<tr>
<td>10-19</td>
<td>362</td>
<td>1.11</td>
</tr>
<tr>
<td>20-29</td>
<td>1433</td>
<td>1.35</td>
</tr>
<tr>
<td>30-39</td>
<td>464</td>
<td>1.59</td>
</tr>
<tr>
<td>40-49</td>
<td>177</td>
<td>1.78</td>
</tr>
<tr>
<td>50-59</td>
<td>36</td>
<td>2.03</td>
</tr>
<tr>
<td>60-69</td>
<td>40</td>
<td>2.33</td>
</tr>
<tr>
<td>70-79</td>
<td>16</td>
<td>4.38</td>
</tr>
<tr>
<td>80-99*</td>
<td>14</td>
<td>4.71</td>
</tr>
<tr>
<td>100</td>
<td>18</td>
<td>7.00</td>
</tr>
</tbody>
</table>

* - the small numbers of cases beyond the 80-yard line suggested a modification of the bin pattern.

To help put the role of field position into perspective, we also regressed points scored against field position. We obtained this simple linear regression equation:

\[
\text{POINTS} = 0.337 + 0.042 \times \text{FIELDPOSITION}
\]
The slope coefficient is significantly different than zero (p<0.001) but the regression's R-squared is only 3.4%. Field position therefore explains very little of the observed variation in points scored on the initial post-kickoff possession.

After a regular kickoff that was successfully fielded by the receiving team, the receiving team, on average, started with the ball near the 26 yard line. More precisely, the mean, median, and standard deviation for starting field position of the receiving team were 26.4, 24, and 11.3 yards. A histogram of field positions is shown in figure 1. The spike at the 20 yard line reflects the influence of the 422 touchbacks. The distribution of field positions is skewed to the right by occasional long run backs, including those returned for a touchdown. For the unusual case (26 events) of the kicking team recovering a fumble after a regular kick, the mean, median, and standard deviation were 75.3, 77, and 6.9 yards (yards from kicking team's goal line).

Figure 1. Starting field position of the receiving team after a regular (not onside) kickoff.

A touchback resulted in the receiving team starting at their 20 yard line (barring a very rare penalty on a touchback kickoff). When the receiving team declined the touchback option the mean starting field position was 24.9 yards (median 22 yards, standard deviation 13.0 yards), which was significantly higher
than the 20 yard line associated with a touchback (p<0.001, one sample t-test). Figure 2 shows a histogram of the field positions after declining a touchback option.

Figure 2. Receiving teams field position after declining the touchback option (running the ball out of their end zone).

3c. Points scored on initial post-kickoff possession

The receiving team scored 1.39 mean points (95% CI: 1.27, 1.51) on the initial post-kickoff possession after a regular kickoff (figure 3). Touchbacks resulted in lower mean points than fielded regular kickoffs but the difference was not statistically significant (1.19 versus 1.43, t-test p=0.12). Similarly, kickoffs that reached the end zone (touchback or run out) were associated with lower mean points than regular kickoffs that did not reach the end zone (1.32 versus 1.44, t-test p=0.33). Declining a touchback option resulted in higher mean points than taking a touchback, but the difference was not statistically significant (1.41 versus 1.19, t-test p=0.23).
Figure 3. Points scored on initial post-kickoff possession after a regular (not onside) kickoff.

Points scored on the initial post-kickoff possession after a surprise onside kickoff ranged from -7 to 8 (negative scores reflect points scored by the kicking team), with a mean of 0.77 (95% CI: -2.40, 3.93; figure 4). This was lower than the mean points scored after a regular kickoff (surprise onside 0.77, regular 1.39 points), but the difference was not significant (p=0.68, t-test). Note that the lower mean points for the surprise onside kickoff is favorable from the kicking team's perspective.
Figure 4. Points scored on initial post-kickoff possession after a surprise onside kickoff.

4. Prospect theory model

We assume a coach's utility function, $U$, over points scored, $x$, of the form:

$$U(x) = \begin{cases} (x)^{\alpha}, & x \geq 0 \\ -\lambda \cdot (-x)^{\alpha}, & x < 0 \end{cases} \tag{1}$$

where $\lambda \geq 1$ and $\alpha > 0$. We define a prospect (or lottery), $P = (p_1, x_1; \ldots; p_n, x_n)$, where the $p_i$ are probabilities and the $x_i$ are points scored outcomes. Observe that a prospect is essentially the probability mass function of a discrete random variable, points scored under a given strategy. We can approximate the probability mass function with an empirical frequency distribution, such as those represented in the histograms of figures 3 and 4. We assume a linear preference functional over a prospect of the form:

$$PT(p_1, x_1; \ldots; p_n, x_n) = \sum p_i \cdot U(x_i) \tag{2}$$
Observe that $U(\cdot)$ is the utility for a given scoring outcome while $PT(\cdot)$ can be interpreted as the utility of a given prospect (or lottery). The linear preference functional, $PT(\cdot)$, is analogous to the expected utility function, $EU(\cdot)$, in expected utility theory. The form of (2) is intentionally simpler than the more general "weighted" form of prospect theory preference functionals (Kobberling and Wakker, 2005). This simplification reduces the number of unknown parameters in our prospect theory model and makes parameter estimation less daunting.

From our kickoff data we can define the four prospects that are involved in the two kickoff decisions (surprise onside versus regular kickoff, decline touchback versus accept touchback). To make comparisons of the utility of a prospect more natural we adopt the convention that positive points scored are favorable to the decision making team under consideration (be it the kicking or receiving team). The prospects, or equivalently the empirical frequency distributions, are displayed in Table 2.

Given any values of parameters $\lambda$ and $\alpha$ we can calculate the preference functionals given in (2). To be consistent with the observed use of these strategies over the 2009 season we must have the following ordering:

1. For surprise onside kickoffs versus regular kickoffs:

\[
PT(\text{regular}) \geq PT(\text{surprise onside}),
\]

since surprise onside kickoffs are used so infrequently.

2. For declining a touchback option versus accepting a touchback:

\[
PT(\text{decline touchback}) \geq PT(\text{accept touchback}),
\]

since touchbacks are so commonly declined. (We acknowledge that Equation (4) depends on how deep the kick was into the endzone. For example, returning the ball out of the endzone when it is more than 3 yards deep could in principle yield fewer points on average than accepting a touch back. However, we ignore these uncommon events for simplicity.)
Table 2. Prospects (empirical frequency distributions) for four kickoff strategies*

<table>
<thead>
<tr>
<th>Kicking team's strategic choices:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular kickoff</td>
</tr>
<tr>
<td>Points</td>
</tr>
<tr>
<td>Probability</td>
</tr>
<tr>
<td>Surprise onside kickoff</td>
</tr>
<tr>
<td>Points</td>
</tr>
<tr>
<td>Probability</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Receiving team's strategic choices:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decline touchback</td>
</tr>
<tr>
<td>Points</td>
</tr>
<tr>
<td>Probability</td>
</tr>
<tr>
<td>Accept touchback</td>
</tr>
<tr>
<td>Points</td>
</tr>
<tr>
<td>Probability</td>
</tr>
</tbody>
</table>

* Scoring outcomes that are favorable to the decision making team are positive

To estimate the model's parameters we start by assuming the most risk neutral and least loss averse preferences for the coaches. That is, $\alpha = 1$ and $\lambda = 1$. That gives:

$PT_{\text{regular}} = -1.389, \quad PT_{\text{surprise onside}} = -0.769$

$PT_{\text{decline touchback}} = 1.411, \quad PT_{\text{accept touchback}} = 1.185$

These parameters do not satisfy the preference functional relationships given in (3) and (4). The observed choices over uncertain kickoff outcomes contradict a hypothesis that coaches have neutral preferences over risk and loss. We consider this an important finding.

Next we decrease alpha and increase lambda until the two preference functional inequalities, (3) and (4), are first satisfied. To operationalize this we...
create a loss function that is minimized subject to the two inequality constraints given above. That is,

\[
\min (1 - \alpha)^2 + (1 - \lambda)^2
\]

s.t. \[ PT \text{ (regular)} \geq PT \text{ (surprise onside)} \]
\[ PT \text{ (decline touchback)} \geq PT \text{ (accept touchback)} \] \hspace{1cm} (5)

Solving this non-linear optimization problem yields parameter estimates:

\[ \alpha = 0.66 \text{ and } \lambda = 1.55. \]

Substituting the parameters into Equation (1), the utility function is depicted in figure 5. It exhibits both modest risk aversion \( (\alpha < 1) \) and loss aversion \( (\lambda > 1) \). To put this utility function into perspective we note that Tversky and Kahneman (1992) estimated values of \( \alpha = 0.88 \) and \( \lambda = 2.25 \) in a controlled experiment of graduate students' decisions over uncertain monetary outcomes. Prospect theory models are notoriously difficult to estimate (Tversky and Kahneman, 1992; Kobberling and Wakker, 2005) so we shouldn't make too much of the difference between model parameters in our work on NFL coaches and Tversky's experiments with graduate students. On the contrary, it is reassuring that our estimated parameter values are indeed plausible.

**Figure 5. Estimated utility function over points scored**
We used points scored on the initial post-kickoff possession as the outcome measure of interest. Other researchers have worked with field position, and then mapped field position to expected points scored (Carter and Machol, 1971; Caroll et al., 1988; Kovash and Levitt, 2009).

Using expected points as an outcome measure in the prospect theory model does not substantially alter our findings (field position to expected points mapping is given in the appendix).

5. Discussion and conclusions

Romer (2006) raised an important question in his paper on 4th down decision making. Are observed patterns of conservative play calling behavior the result of imperfect optimization by coaches, or do these patterns arise because coaches are risk averse? It is easy to conflate imperfect optimization and non-neutral risk preferences. In both cases the quantitative football analyst observes "irrational" decision making. As Romer (2006) pointed out, there is little experimental evidence to answer this question.

Many researchers explicitly or implicitly assume neutral risk preferences in their analysis. It improves the tractability of models, so the practice is appealing. Kovash and Levitt (2009), for example, used a strictly competitive game theory (minimax) model to study the pass versus run decision in the NFL. A minimax model implicitly assumes risk neutrality since a game with more than two simple win-lose outcomes ceases to be zero sum if non-linear preference functionals are permitted (Binmore, 2007). Kovash and Levitt's finding that coaches play suboptimally, given an implicit assumption of risk neutrality, could therefore be interpreted in another light; coaches might be playing optimally if they possess risk averse preferences.

A minority of football researchers have explicitly assumed that football coaches are risk averse. Rockerbie (2008), for example, used analogies to financial markets in his research on the "passing premium puzzle" (a concept first described by Alamar, 2006). In effect, he assumed that NFL coaches are risk averse over uncertain field positions since the general population is typically risk averse over uncertain monetary outcomes. We are not convinced by this argument, but remain sympathetic to the general idea of non-neutral risk preferences.

In this paper we studied the use of kickoff strategies from the 2009 NFL season. Coaching choices in this facet of football strategy were found to be inconsistent with the notion that coaches have neutral preferences over risk and loss. Estimation of a postulated prospect theory model yielded parameter estimates indicative of both modest risk aversion over gains and loss aversion over losses. Coaches display diminishing sensitivity to changes in scoring.
outcomes as they move further from a reference point (zero), and for scoring gains and losses of equal magnitude they suffer more from a loss than they enjoy from a gain. We view this research as a humble first attempt at answering the research question originally posed by Romer (2006). We suspect that deviations of coaches' decisions from seemingly optimal play can be decomposed into two components: that due to non-neutral risk preferences and that due to imperfect optimization. Estimating the relative contribution of the two components is not straightforward. Our research did not address this question. Finally, we wonder if it makes sense for NFL coaches to be risk averse. Over time, shouldn't risk neutral coaches experience greater success and drive risk averse coaches out of the NFL market? These are interesting issues for further research.

We acknowledge that in practice, coaches may not make the decision whether to decline a touchback. For example, if a return man catches the ball and has a split second, in the course of play, may decide whether to run the ball out of the endzone or not. For simplicity and tractability, this paper focuses on the cases where coaches make all the decisions. In the future we could study the cases where players make decisions in the course of play.

References


