Estimating the Value of "Going For It" (When No One Does)

Christopher P. Adams[1]
Federal Trade Commission

December 12, 2006

Abstract

This paper estimates success rates and the value of "going for it" on fourth down using three alternative approaches. First, the paper uses actual fourth down attempts from all regular season NFL games played in 1998, 1999 and 2000. It uses early quarter data and functional form assumptions to estimate success rates at various to go distances. It also uses all fourth down attempts and a Heckman selection model to account for the selection bias in the data. Second, the paper estimates success rates at various positions and to go distances from simulated data using Madden NFL 07. Third, the paper estimates a game theoretic structural model of fourth down attempts using actual third down data and estimated expected points. Romer claims to demonstrate that American football teams systematically err in their choice of whether to punt or "go for it" on fourth down, and uses those results to argue that firms do not maximize profits. The analysis below calls into question whether Romer indeed successfully demonstrated that football teams punt too often, which in turn casts doubt on the more general conclusions he drew from his results.

[1] Email: cadams@ftc.gov. Thanks to David Romer and Mario Lopez for providing me with the data used in Romer (2006). Thanks to David Schmidt for suggestions and providing data on point spreads and over-unders. Thanks to C.J. Adams for helping with running simulations on Madden NFL 07. Thanks to Eric Durbin, Dan Hanner, Dan Hosken, Jesse Leary, Rob McMillan, Joe Mulholland, Michael Salinger, Dave Vanness, Abe Wilkelgren and Justin Wolfers for comments and suggestions. This paper does not represent the views of the Federal Trade Commission or any individual Commissioner. Any errors are my own.

1. Introduction

Rationality and optimizing behavior are often a central assumption in economic models. Are economic agents rational? Can economic agents solve complicated optimization problems? A body of literature has grown up testing this assumption in the laboratory (Camerer (2003) for example). There has been less work testing this assumption in the field. One recent exception is Levitt (2006), which argues that even an MIT trained Ph.D. economist is unable to choose bagel prices optimally. In a similar vain, Romer (2006) argues that NFL coaches fail to optimally solve dynamic programming problems. Romer (2006) argues that NFL coaches fail to go for it enough on fourth down and in particular, these coaches leave expected points on the table. If true, Romer's result suggests that economic actors may not be the rational optimizers often assumed in economic models. In particular, Romer argues the results suggest firms do not maximize profits. This paper takes a closer a look at the analysis and presents results that suggest NFL coaches behave in a way that is consistent with optimizing behavior and thus casts doubt on the conclusion that firms do not maximize profits.

One concern with empirical analysis of going for it on fourth down is that we almost never observe teams going for it on fourth down. To overcome the lack of data, Romer (2006) assumes that success rates on third down can proxy for success rates on fourth down. This paper takes a closer look at this assumption and presents results from three alternative strategies for estimating fourth down success rates. Two of the strategies lead to estimated success rates consistent with Romer (2006), the third approach (the structural model) leads to estimates that are quite different from Romer (2006). Moreover, the structural estimates suggest that NFL coaches may act as if they can solve complicated dynamic programming problems.

The paper presents three alternative strategies for estimating success rates on fourth down. First, the paper uses actual fourth down attempts. The paper presents alternative approaches for identifying success rates at various to go distances and field positions from actual fourth downs. The paper uses data from early in the game and assumes there is little selection problem in regards to observed fourth down attempts. However, this data has few observations at the longer to go distances so functional form assumptions are needed for identification. The paper also uses all the fourth down attempts and a selection model (a "Heckit") to account for selection into fourth down and therefore estimates "adjusted" success rates. Second, the paper estimates success rates at various field positions and to go distances using simulated data. Simulations are run on Madden NFL 07 using random assignments of teams into field positions and to go distances. Offensive play calls are also selected at random but the success rates are calculated using "optimal play calls". Third, the paper estimates a game theoretic structural model of third down situations and fourth down attempts. Third down data on play calls and success rates and information on expected points from Romer (2006) is used to estimate parameters of the model. Success rates on fourth down from various field positions and to go distances are calculated assuming a mixed strategy equilibrium using parameter estimates and expected points.

The results from using actual fourth down data and the results from the simulated fourth down data are generally consistent with the results presented in Romer (2006). Using third downs as a proxy for fourth down, Romer (2006) determines the value of going for it versus kicking (kicking a field goal or punting) from various field positions and to go distances. Romer's results suggest that NFL coaches should go for it much more often than they do, particularly in their own half of the field. Results from the structural model are not consistent with Romer's findings. In particular, the success rates on fourth down from the structural model suggest teams should generally not go for it in their own half, a result consistent with observed behavior. It should be noted that this prediction is based upon the same model of decision making as Romer (2006) and the same expected points for various first down positions and kicks as Romer (2006). The different predictions are due to the different estimated success rates from the structural model. Moreover, the results contradict the statement in Romer (2006) that equilibrium outcomes on fourth down would not differ from equilibrium outcomes on third down.

During a typical NFL game the coach or offensive coordinator faces the repeated question of whether to go for first down on fourth down, to punt the ball away or to kick a field goal. In most cases, the coach chooses to either punt or kick a field goal. The choice between going for it and kicking is a classic dynamic decision making problem under uncertainty (Rust (1986) for example). A successful pass or run will advance the ball down the field and increase the probability of a score and decrease the probability that the defense scores. An unsuccessful pass or run will lead to a turnover on downs giving the opponent the ball at the same position. On the other hand the team could punt and have the opponent start at a position that is significantly further away from their own goal. Thus, relative to a turnover on downs a punt reduces the likelihood of the opponent scoring. If close enough, a field goal kick will lead to 3 points. While 3 points is not the 7 points that is generally received from a touch down it may be better than nothing.

Do NFL coaches make this choice optimally? Romer (2006) suggest that one way to determine the answer is to reduce the problem to a problem of maximizing expected points from each choice and compare predicted choices to actual choices. While winning is obviously of paramount importance, Romer argues maximizing expected points is probably a good approximation of behavior in the first quarter of the game. To predict optimal choices it is necessary to calculate the expected points a team would receive on the possession given the current field position and the expected points from punting and kicking. The difficulty is that are many states of the world to consider. At the teams own 30 yard line a pass on fourth down may lead to a touch down, a completion moving the team to the 50 yard line, an incomplete pass and turnover on downs, an interception, an interception returned for a touch down, a five yard completion, etc etc. To simplify matters Romer (2006) breaks the state space into first downs for each team at each yard line on the field and various scores (something over 200 states). The paper assumes there is a Markov process that moves the team from one state to every other one of these states. Assuming this stationary process and that there is a single representative team, Romer (2006) estimates the marginal effect of the state on the team's score. For example the expected points from the team's own 10 is about 0, the expected points from the 50 are about 2 and the expected points from the opponent's 10 is about 5. Romer (2006) similarly estimates expected points from punts and field goals from various field positions. Thus, with these estimates all that is needed to predict optimal choices is an estimate of the success rate on fourth down.

Romer (2006) considers using actual fourth down data (and does use it to compare results from third and fourth down data) but notes that there are two problems with the data.[2] First there are few observations and second the "times when teams choose to go for it may be unusual: the teams may know they are particularly likely to succeed, or they may be

[2] Romer (2006) states that he uses all the Fourth Down attempts accept for the last two minutes of each half and overtime. He says he also accounts for point differential and point spread (although it is not clear from the description how this is done). He finds that success rates from actual Fourth Down data is one percentage point lower than success rates from the Third Down data and that this difference would not affect the analysis.

desperate." (p. 357) In the first three quarters teams almost only go for it when there is less than one yard to go.[3] One way to estimate success rates at the longer to go distances is to use functional form assumptions to project into the area that is not observed. A second approach is to use data from the fourth quarter and overtime because in this data we observe fourth down attempts at the longer to go distances. The problem with this "late game" data is that there may be significant selection problems. A standard method for accounting for this "selection" into going for it is to explicitly model the selection and adjust the estimates accordingly (Greene (2003)). This paper uses a "Heck Probit" which allows the error on the selection into going for it to be correlated with the error on the success probability. To estimate the selection into going for it or kicking (kicking a field goal or punting) the paper uses a large number of observable characteristics of the team and the situation including team dummies (offense and defense), point spreads, over-under, home field, quarter, position on the field, to go distance, time to go in the quarter, date dummies and various non-linear terms. In the reported estimates success rate is a function of the to go distance and the to go distance squared.

The results confirm the statement in Romer (2006) that it makes little difference whether third down or fourth down data is used. The results from using fourth down attempts from the first three quarters of the game suggest that using third down data may give biased results, but the direction of the bias is dependent on the to go distance. When all the fourth down attempts are used and the success rates are adjusted by the selection model the success rates from the fourth down (almost) always lie below the success rates from the third down data. However, at most the difference is less than 1.5 percentage points and the two rates are not statistically significantly different.

Still, caution should be used when interpreting results from the Heckit model. What we would like to know is the probability of success of going for it for the team in the situation in which that team in that situation explicitly chose not to go for it. While we may be wary of assuming that NFL coaches are rational, it is unlikely that these same coaches make such decisions randomly. Presumably there is information that is available to the coach about the likely success rate that is unavailable to the statistician. While, the Heckit

[3] One problem with the data is that it is not possible to distinguish between 4th-and-inches and 4th-and-1, a difference which my own observations suggests is of great importance.

model accounts for this selection it does so using particular parametric and functional form assumptions.[4]

The paper's second strategy for estimating success rates on unobserved fourth downs is to simulate them. The paper takes advantage of Madden NFL 07 to simulate outcomes of fourth down attempts given various situations and various teams. The paper presents results from a couple of different scenarios including random selection of pass or run, optimal selection of pass or run, computer controlled defensive strategy and optimal defensive strategy. The results suggest that success rates on fourth down may be lower than success rates on third down particularly in longer yardage plays (more than 1 yard). Allowing, the offensive to use runs on short yardage plays gives success rates on fourth down that are actually higher than those from third down. However, if the defense plays a "goalline" defense with 1 yard to go, offensive success rates are significantly reduced and are below the third down plays.[5] The advantage of simulating the plays is that it is possible to observe success rates for situations in which teams would not normally go for it on fourth down. The obvious concern with this approach is that Madden NFL 07 is a black box. It is not clear what is driving the results. What type of data is used to determine the success of a particular play? Is it aggregate play level data such as Romer (2006) uses or is it player level data or some combination of the two?

The paper's third strategy for estimating success rates from the unobserved fourth down data is to use the third down data to estimate parameters of a game theoretic model of play selection on third down. The structural parameter estimates are then fed into a game theoretic model of play selection on fourth down. Success rates are estimated from the parameter estimates assuming equilibrium play on fourth down. This approach follows the recent tradition of using equilibrium behavior in a game theoretic model to identify the underlying parameters of the model (Berry et al (1995), Bajari et al (2006) for example). In this case the third down game is used to identify the probability of success conditional on the actions chosen by the defense and the offense. The paper shows that with these underlying conditional probabilities it is possible to identify the overall success rate from the

[4] See Manski (1995) for alternative methods. Unfortunately, a bounds approach would posit a level of rationality that assumes the result.

[5] The computer only plays a goal line defense on fourth and inches. When the offense has 1 yard to go, the computer plays a more standard defense. It is possible to do better by overriding the computer and playing a goalline defense when the offense has 1 yard to go. The standard goalline defense used in Madden NFL 07 is one in which there are 7 down linemen, two line backers and two defensive backs.

equilibrium of the fourth down game.[6] The paper uses a two-step structural estimation procedure (Bajari et al (2006)). In the first step the probability the Offense plays Pass conditional on observed characteristics and the probabilities of success conditional on the offensive play call and observed characteristics are estimated using a standard probit model. In the second step, the fourth down success rates conditional on various characteristics are determined by placing the estimated play call and success rate probabilities into the model. Standard errors are calculated using a bootstrap procedure.

Romer (2006) considers the issue of how the "game" changes between third and forth down.

> Relative payoffs to different outcomes are different on the two downs. In particular, the benefit to a long gain relative to just making first down is smaller on fourth down. As a result, both the offense and the defense will behave differently: the offense will be willing to lower its chances of making a long gain in order to increase its chances of making a first down and the defense will be willing to do the reverse. (p. 356).

The paper then goes on to explain why the Nash equilibrium outcome would not change despite these changes in payoffs (although the paper never uses the term Nash equilibrium),[7]

> Since it seems unlikely that the defense has substantially more influence than the offense on the distribution of outcomes, it follows that the use of third downs is unlikely to lead to substantial overestimates of the value of going for it. (p. 356)

The author also states that the difference in payoffs is not large and for this reason (also) the outcomes between the third down game and the fourth down game would not be different.

This paper directly models the game on third down and the game on fourth down. It uses observed success rates and pass/run ratios to estimate the parameters of the model and then considers how the Nash equilibrium outcome on fourth down changes when the payoffs change. The results suggest that moving from third down to fourth down *significantly* changes the probability of success on fourth down.

The results from the actual data and the simulated data are similar to the third down data used by Romer (2006). The results from the structural model are quite different. This

---

[6] The assumed difference between third down and fourth down is the decrease in expected points to the offense from failure.

[7] In personal correspondence with the author, Romer did say that he was referring to a Nash equilibrium.

paper compares the predictions on when to go for it from Romer (2006) and the structural model with actual behavior. Romer (2006) predicts teams should always go for it on their own 20 with 3 yards to go. In actual fact teams almost never go for it in such a situation. The structural model predicts such low success rates in this situation that the team should not go for it. While Romer (2006) argues NFL coaches are not rational dynamic optimizers, this paper suggests that they may be or at least act as if they are.

The paper continues as follows. Section 2 analyzes the probability of going for it on fourth down. Section 3 estimates success rates from actual fourth down data. Section 4 presents success rates from simulating fourth down attempts in Madden NFL 07. Section 5 presents success rate using estimates from a structural game theoretic model of third down plays. Section 6 discusses the implications of the success rate estimates for the value of going for it on fourth down. Section 7 concludes.

2. Going For It on Fourth Down

This section analyzes the determinants of "going for it" in the NFL. The purpose of this section is twofold. First, the section aims to get a better handle on why coaches choose to go for it. Second, the section develops a model that will be used in the next section to account for selection into going for it. If there is a takeaway from this section it is that teams almost only ever go for it with less than a yard to go.

The paper uses the same data as Romer (2006) which includes all the regular season NFL games played in seasons 1998, 1999 and 2000. The data includes information on the type of play chosen on fourth down including punt attempts, field goal attempts and attempts to "go for it". The data also includes information on the situation including team with the ball, opponent, yards to go, position on the field, quarter, score, and minutes remaining. The data includes time information such as date and the "week" of the game. Romer's data includes point spreads and I have augmented the data with a measure of over-under.[8]

[8] The point spread is a particular bookie's measure of estimated average point differential between the two teams and over-under is a bookie's measure of the estimated average number of total points scored. It is not clear how much these differ across bookies and it not clear whether they are a measure of market beliefs or the

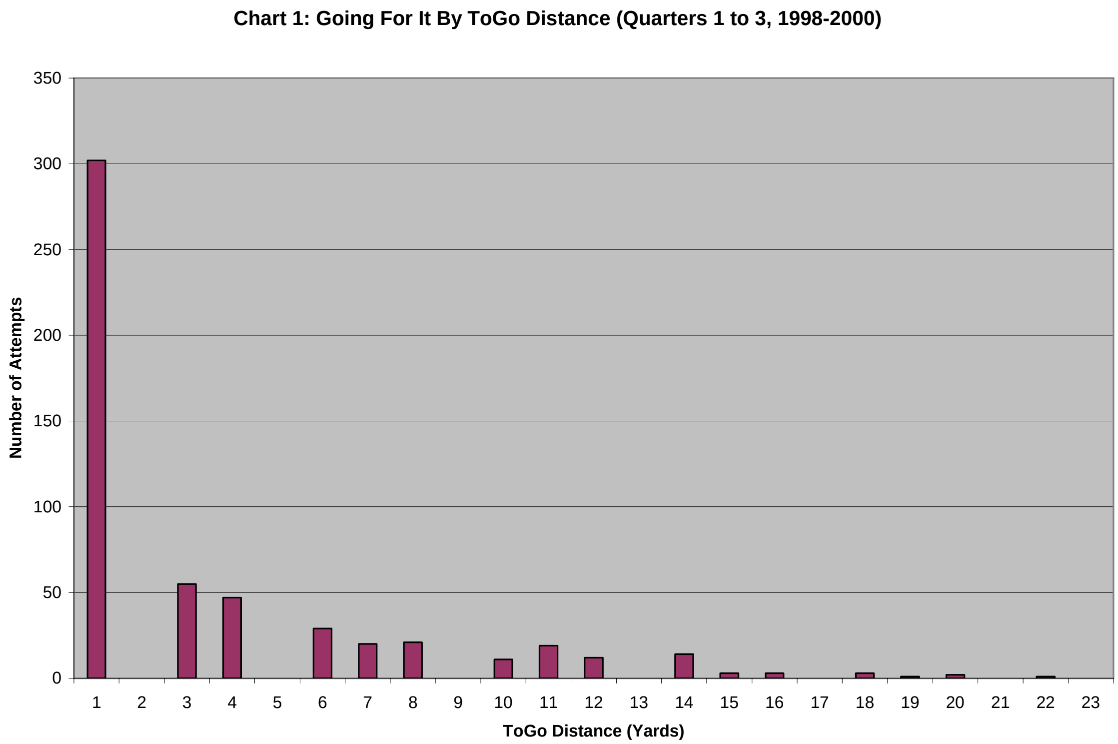

Chart 1 presents the actual number of attempts to go for it by the number of to go yards left for first down or goal. This graph includes all games but only attempts made in the first three quarters. One quirky issue in the data is that the information provider codes all to go distances that are less than 1.5 yards as 1 yard. Romer recodes this data as .75 yards. Importantly, no distinction is made between 4th-and-1 and 4th-and-inches even though anecdotal evidence suggests the difference is great. Even so, the graph makes it pretty clear that the vast majority of cases in which a team decides to go for it on fourth down involve 1 yard or less to go (at least in the first three quarters).

actual state of the world (see Adams (2006) for discussion of the information that can be learned from prediction markets).

Table 1: Probit for Going For It on 4th Down (All attempts 1998-2000)

Variable	*dF/dx*	*Std. Err.*	*dF/dx*	*Std. Err.*	*dF/dx*	*Std. Err.*
0 to 1.5 Yards To Go (Dummy)	0.23	0.02	0.21	0.02	0.22	0.02
To Go (10xYards)	-0.03	0.01	-0.03	0.00	-0.03	0.00
10 Yard Line or Less (Dummy)	-0.08	0.01	-0.03	0.00	-0.03	0.00
10 to 20 Yard Line (Dummy)	-0.07	0.01	-0.03	0.00	-0.03	0.00
20 to 30 Yard Line (Dummy)	-0.08	0.01	-0.04	0.00	-0.04	0.00
30 to 40 Yard Line (Dummy)	-0.07	0.01	-0.04	0.00	-0.03	0.00
40 to 50 Yard Line (Dummy)	-0.05	0.01	-0.03	0.00	-0.03	0.00
50 to 60 Yard Line (Dummy)	0.00	0.01	0.00	0.01	0.00	0.01
60 to 70 Yard Line (Dummy)	0.13	0.02	0.12	0.02	0.12	0.02
70 to 80 Yard Line (Dummy)	0.03	0.01	0.03	0.01	0.03	0.01
80 to 90 Yard Line (Dummy)	-0.02	0.01	-0.01	0.00	-0.01	0.00
Points Up (10xPoints)	-		-0.02	0.00	-0.02	0.00
Points Up Late (10xPoints)	-		-0.02	0.00	-0.02	0.00
Time To Go in Quarter (10xMinutes)	-		-0.02	0.00	-0.02	0.00
Time To Go Late (10xMinutes)	-		-0.08	0.01	-0.08	0.01
First Quarter (Dummy)	-		-0.08	0.00	-0.08	0.00
Second Quarter (Dummy)	-		-0.09	0.01	-0.08	0.01
Third Quarter (Dummy)	-		-0.07	0.00	-0.07	0.00
Points Favored (100xPoints)	-		-0.02	0.03	-0.02	0.03
Over-under (100xPoints)	-		0.08	0.03	0.13	0.05
Home (Dummy)	-		0.00	0.00	0.00	0.00
Week (10xIntergers 1 to 17)	-		-		0.01	0.00
1998 Season (Dummy)	-		-		0.01	0.00
1999 Season (Dummy)	-		-		0.00	0.00
Team Dummies	No		No		Yes	
Team Defense Dummies	No		No		Yes	
Log Likelihood	-3,396		-2,325		-2,267	
Adj R^2	0.18		0.43		0.45	
N	11,385		11,358		11,358	

Notes
1. It does not include plays with penalties or fakes.
2. Late refers to the 4th quarter or beyond

Table 1 presents the probit results for going for it on fourth down for all attempts for all teams from 1998 to 2000. The analysis compares going for it to *either* punting or kicking a field goal. The analysis does not include plays in which there was a penalty or a fake. The table presents result from three probit regressions. The first column in each case is the change in percentage points given a small change in the variable. The second column in each case is the standard error on the derivative. The three models are presented in the order of the number of explanatory variables included. The first model includes only measures of the to go distance and the field position. The second model includes measures of the situation including point differential, time, quarter and point spread. The third model includes time and team dummy variables.

As Chart 1 suggests, being 1.5 yards or less to the first down or the goal line has an extremely large effect on the probability of going for it. The overall probability of going for it is just under 12% and at the average of the variables the probability of going for it is 3%. At the average of the variables, being 1.5 yards or less to the first down increases the probability of going for it by 22 percentage points! The results also suggest that there is a substantial increase in the probability of going for it when the ball between the opponent's 40 and 30 yard lines. The probability of going for it when the ball is between the opponent's 40 and 30 is 13 percentage points higher than the probability of going for it when the ball is between the opponent's 10 and the goal line. The probability of going for it is between 5 and 8 percentage points less when the ball is in one of the 10 yard zones in the team's own half relative to 10 or less and goal to go. Observable factors about the state of the game have some impact on the probability of going for it. Teams that are down are more likely to go for it. Another 10 points down increases the probability of going for it by 2 percentage points. Teams are more likely to go for it with less time remaining in the quarter, particularly late in the game. In the fourth quarter, two and half minutes less time remaining increases the probability of going for it by 2 percentage points. Teams are also much more likely to go for it in the fourth quarter and overtime when they are underdogs and when the game is expected to be high scoring. There is some evidence that the team is more likely to go for it late in the season. Surprisingly the home team does not seem to be more likely to go for it than the away team.

As stated above, short yardage has a huge effect on the probability of going for it. Unfortunately, it is not possible to determine whether for the majority of the short yardage situations involve to go distances that are only a couple of inches. My own observation suggests that teams do go for it on 4th-and-inches and the offense runs a quarterback sneak. Teams also go for it is when the ball is between the opponent's 40 and 30 yard line. This field position is something of a black hole because it is often too far for an accurate field goal and the value of a punt can be limited particularly if there ends up being a touch back to the opponents 20 yard line.[9]

A number of the results suggest that coaches are not simply considering the expected value of going for it on fourth down. It seems that coaches are also interested in the variance and are much more willing to "take a chance" in certain situations. In particular, teams are more likely to go for it when they are down, when it is late in the quarter, when it is late in the game, when they are underdogs and when the expected aggregate score is high.[10] There is also some suggestion that the team is more likely to go for it late in the season. When a team is down by large number of points, expected to lose by a large number of points or when they are playing in a high scoring game, the value of a touchdown relative to the value of a field goal or punt may be substantially higher.

3. Estimating Success Rates from Fourth Down Data

Romer (2006) uses the success rates from third down data to determine whether NFL coaches go for it enough on fourth down. One major concern with this approach is that fourth down is not the same as third down. Romer (2006) acknowledges this problem but argues that it is not possible to use success rates from actual fourth downs because of the lack of data and selection problems with the data that is available.[11] Selection problems have become one of the classic identification problems in economics. What we would like to do is randomly assign teams into fourth down plays and observe the outcomes given different field and game situations. In this way we can measure the "treatment effect" of going for it with 4 yards to go relative to going for it with inches to go. By how much does the extra 4

[9] Thanks to David Meyer for pointing out this possibility.
[10] Thanks to Dan Hanner for suggesting over-under as a measure and thanks to David Schmidt for supplying the data. Over-under is the expected aggregate score of both teams.
[11] Romer also does some comparisons between third and fourth down success rates although exactly what is done is not clear from the description in the paper.

(or slightly less) yards increase the probability of success on fourth down. The problem is that, as Chart 1 and Table 1 suggest, coaches are substantially more likely to go for it with less than a yard to go then they are when there is 4 yards to go. This section presents results from two ways around the problem without resorting to using third down data. The first method uses a functional form assumption to project out what would happen with 4 yards to go using fourth down attempts from the first three quarters. The second method uses attempts from all quarters and a Heckman selection model to adjust the success rates to account for selection into going for it.

The two methods have their advantages and disadvantages. In the first, the data is from the first three quarters and so it is less likely that coaches will be "taking chances". The problem is that the identification of the treatment effect is coming from the functional form assumption because there is little data to calculate success rates for longer to go distances. The advantage of the second approach is that we do actually observe coaches going for it with 4 yards to go. The disadvantage is that this is not a random selection of teams into situations and while the Heckman model accounts for this selection, it is still identifying the treatment effect from distributional and functional form assumptions.

Chart 2: Probability of Success by To Go Distance (Probit Models 1998-2000)

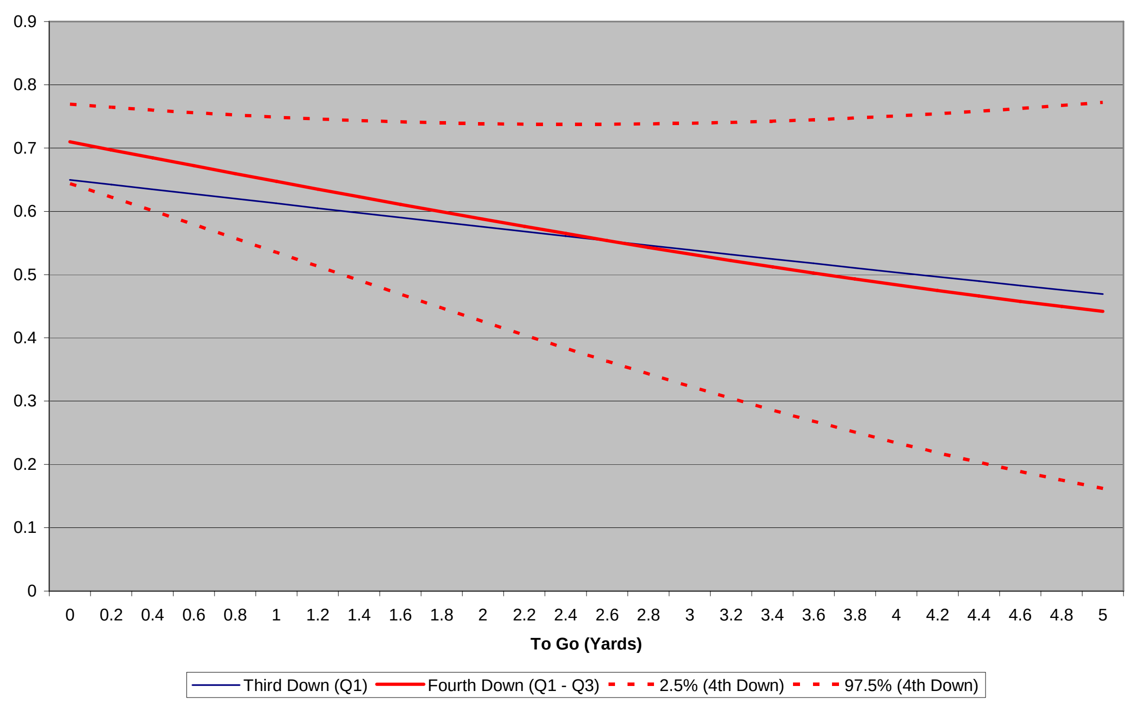

Chart 2 presents a comparison between estimated success rates from actual fourth down data with estimated success rates from third down data. Both lines are calculated from a simple probit model where the mean success rate is a quadratic function of the to go distance. The dotted lines give some sense of the statistical significance of the fourth down estimates. The third down probit model is estimated using data from the first quarter while the fourth down probit model is estimated using the first three quarters. The third down model is estimated with almost 5,000 observations and the fourth down model is estimated with only 544 observations.[12]

The chart suggests that using third down data may produce biased estimates on the value of going for it, although the direction of the bias depends on the to go distance and the third down line always lies within the 95% confidence interval for the estimated model. Taking the point estimates seriously, the chart suggests that it is only after 2.6 yards that using third down data may lead to over estimating the value of going for it. The graph also suggests that using third down data under estimates the treatment effect of moving from 4th-and-inches to 4th-and-4. In the fourth down data the success rate fall much more quickly as the to go distance increases.[13]

One concern with this approach is that we do not have a very accurate measure of the actual to go distance for short yardage situations. In the third down data the to go distance is measured in the integers. In the fourth down data Romer codes all distances that are less than 1.5 yards as 0.75 yards. Other than that, all fourth down to go distances are in the integers. The raw data on fourth down does not distinguish the actual distance for all distances less than 1.5 yards. It is possible to believe that success rates are substantially higher at 4th-and-inches than they are at 4th-and-1 and thus by aggregating the results the treatment effect of moving from 4th-and-inches to 4th-and-4 may be biased downwards.

[12] In unreported results the selection into going for it on fourth down seems to be quite similar for the first three quarters.
[13] The 95% confidence intervals from the third down and fourth down models overlap for the coefficients on the to go distance.

Chart 3: Probability of Success by To Go Distance (Heckit/Probit 1998-2000)

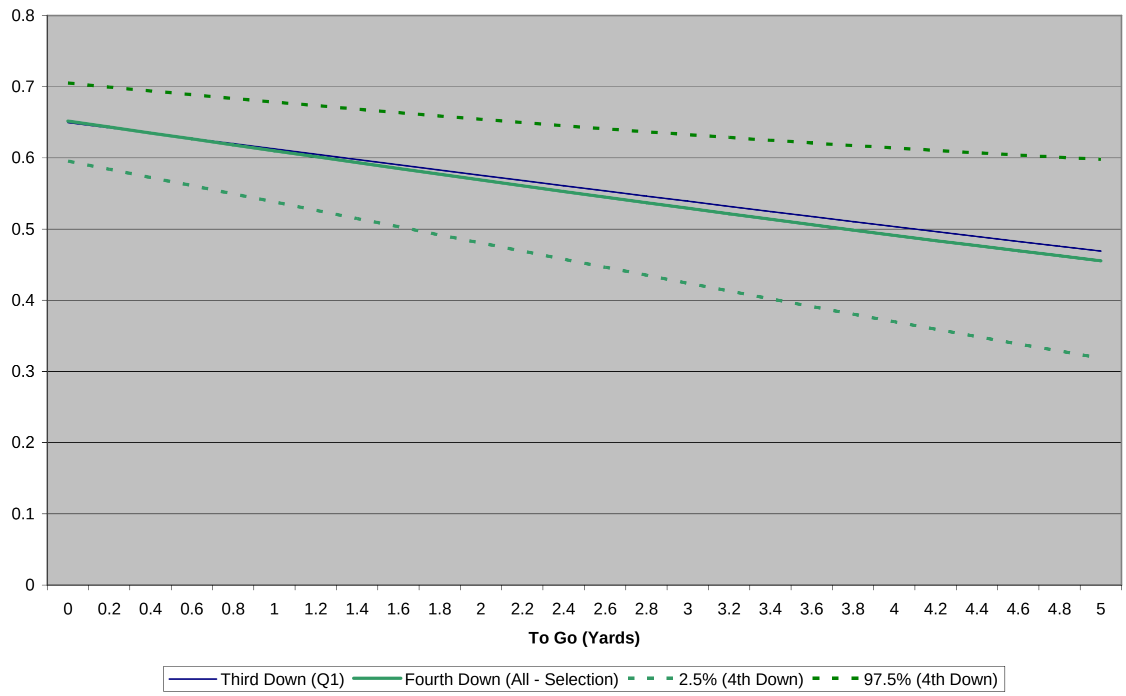

Chart 3 presents the success rates by the to go distance based on the estimates from the selection model and the estimates from the third down data. The results show the success rates from the selection model are almost always lower than the success rates from the third down data. However, the two are not statistically or economically significantly different. The difference increases to at most 2 percentage points with 5 yards to go. According to Romer (2006) a difference of 1 percentage point is not enough to change the main results regarding the value of going for it on fourth down. The Heckit results are presented in the appendix.

This result suggests that Romer (2006) is correct that it makes little difference whether success rates are measured using third down or fourth down data. The advantage of using the selection model is that the treatment effect of a larger to go distance is measured using actual observations from the larger to go distances. The model accounts for the concern that there is not a random selection into going for it on fourth down. There are some concerns with this model because the adjustments made to the success rate estimates come from parametric and distributional assumptions.

4. Simulating Fourth Downs

The concern with using fourth down data is either there is not enough data on the success rates for going for it at larger to go distances or if there is data there is a substantial sample selection problem. This section presents results from simulating going for it on fourth down using Madden NFL 07. Madden NFL is the best selling football simulation game. In part, the game uses individual player ratings and statistics to determine the simulation results from various plays.[14] Madden NFL is arguably the next best thing to actually having a team randomly choose to go for it on fourth down.

The objective of the simulations was to "randomly" vary the situations and teams to see how these factors affected average success rates. For each simulation the field position was varied by 10 yards from the team's own 10 to their opponent's 10 and the to go distance was varied by 1 yard from 5 to inches. The home and away teams were also varied. Unfortunately, it was not possible to simply set up the situation and have the computer play both sides of the ball.[15] The reason – like actual NFL coaches the computer does not go for it very often on fourth down. Therefore, human intervention was required to have the team go for it. The play was assigned by choosing one of three formations, generally either an I, a split back or a single back, and choosing "0" or "X" out of each formation. In general, "O" is a pass and "X" is a run.[16] For the most part the simulations were run in "coach" mode which allows the human to call the plays and do the snap but then the computer takes over from there. In general the computer played defense including making the play calls. In a subset of cases the defense was forced to play a goalline defense irrespective of the situation.

[14] See for example http://www.ugo.com/channels/games/features/maddennfl_2003/.

[15] The simulations were run from "practice mode" which allows all aspects of the situation to be adjusted including teams, to go distance, field position, time on clock etc.

[16] For most teams runs from the split back or single back formations are draws (or fake passes) and runs from the I are dives. Similarly passes from the I are a play action pass (or fake runs).

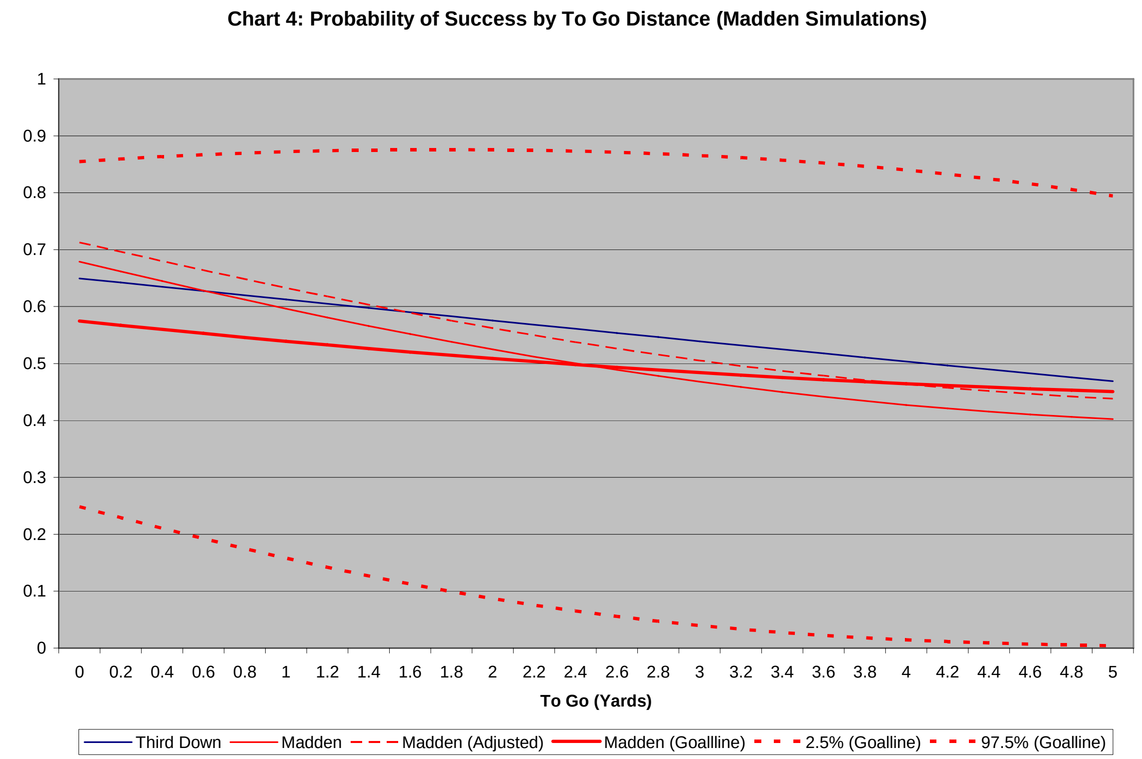

Chart 4 presents the results from estimated success rates based on 535 simulated fourth down attempts. The line denoted "Madden" is based a probit model which only accounts for the to go distance. The line denoted "Madden Adjusted" is based on a probit model which includes a dummy which is 1 if the play was a run with 3 yards or less to go or a pass with more than 3 yards to go and 0 otherwise. The line denoted "Madden Goalline" includes this dummy and a dummy which is 1 if the defense is playing goalline with 1 yard or less to go and 0 otherwise. The "Third Down" line is the same as presented in the previous charts.

The line denoted Madden Goalline suggests success rates from fourth down could be 8 percentage points lower than those predicted by using the third down data. These results also suggest that the treatment effect of moving from 4^{th}-and-inches to 4^{th}-and-4 may be smaller than what would be suggested by the third down data. This said, none of the simulation results are statistically significantly different from each other or the third down results.

My preferred simulation result is the one labeled goalline as this one allows the offense and defense to have the optimal play given the to go distance. In general the simulation results are based on random selection into situations and play calls. A dummy for pass or run makes no difference to the success rates in the probit model, but adding a dummy which 1 for run on short yardage and pass on long yardage is statistically significantly different from 0 and economically significant (which you can see by comparing the Madden line to the Madden Adjusted line). This dummy is optimized when the cutoff for short yardage is the three yard line. Similarly, I experimented with having the computer call the defense and forcing the defense to play goalline. Again, a dummy for goalline was not significant but a dummy for goalline on short yardage and computer on long yardage was statistically and economically significant. The cutoff was optimally (for the defense) set at the one yard line.

The fact that I can improve on the computer generated defense by forcing it to play goalline in short yardage situations suggests a concern with the simulation. In particular, it may be optimal for the offense to choose a particular play against the goalline defense in short yardage situations, for example a quarterback sneak or a quick pass.

5. Structural Estimates from Third Down Data

This section presents results using the third down data. The difference between the results here and those presented in Romer (2006) is that here the third down data is used to estimate parameters of structural model of the situation. The success rates from fourth down are then calculated by making appropriate changes to the model. In particular the relative value of a pass play is decreased when the situation changes from third down to fourth down. This change in the payoffs changes the equilibrium strategies and the equilibrium outcome. Romer (2006) acknowledges that these changes could occur but argues that equilibrium success rates in the third down situation would be not be very different from equilibrium success rates in the fourth down situation. Here, I determine the new fourth down equilibrium and calculate the success rates. The results suggest the difference is significant.

Consider the following zero-sum game. The Offense and the Defense can simultaneously choose between two actions, Run and Pass. The probabilities of success (for the Offense) from each combination of actions are denoted:

Pr(Success | Offense plays Pass, Defense plays Pass, X) = $a(X)$

Pr(Success | Offense plays Pass, Defense plays Run, X) = $b(X)$

Pr(Success | Offense plays Run, Defense plays Pass, X) = $b(X)$

Pr(Success | Offense plays Run, Defense plays Run, X) = $c(X)$

where X is the vector of observable characteristics (including field position and to go distance). Note that we assume success rate is the same on the "off-diagonal". The third down payoff matrix is:

		Defense	
		Pass ($q_3(X)$)	*Run* ($1 - q_3(X)$)
Offense	*Pass* ($p_3(X)$)	$a(X)(f(P+15) - g(P))$	$b(X)(f(P+15) - g(P))$
	Run ($1-p_3(X)$)	$b(X)(f(P+T) - g(P))$	$c(X)(f(P+T) - g(P))$

where $f(P)$ is the expected number of points from first down at a particular position P, $g(P)$ is the expected number of points from kicking (punting or kicking a field goal), T is the to go distance. The values $f(.)$ and $g(.)$ are based on results presented in Romer (2006). Following Romer (2006) expected points do not vary except with field position. As this is a zero-sum game the payoffs are those to the Offense, while the Defense payoffs are simply the negative. The relative expected payoff from Pass is the success rate multiplied the expected number of points of being 15 yards down field less the (net) payoff from a kick.[17] It is assumed that the team that fails on third down will punt or kick, whichever has the higher expected points. The expected payoff from Run is the success rate multiplied by the expected number of points at the same position plus the to go distance (T) less the payoff from kicking.[18]

[17] The assumption that a pass gains 15 yards is arbitrary but consistent with the discussion in Romer (2006). The important difference between "Pass" and "Run" is that the "Pass" is a play that is expected to move the ball significantly further down the field.

[18] Again the assumption that a successful run only gets to the first down line is somewhat arbitrary.

The fourth down payoff matrix:

		Defense	
		Pass ($q_4(X)$)	*Run (1-$q_4(X)$)*
Offense	*Pass ($p_4(X)$)*	$a(X)(f(P+15) + f(100 - P))$	$b(X)(f(P+15) + f(100 - P))$
	Run (1-$p_4(X)$)	$b(X)(f(P+T) + f(100 - P))$	$c(X)(f(P+T) + f(100 - P))$

The payoff matrix is almost the same as for third down. The difference is that instead of getting *g(P)* if the team fails the team gets a net payoff of *-f(100-P)* which is the net expected points from turning the ball over on downs at position *P*.

This is obviously a simplified version of reality. There are a number of important assumptions. First, only two actions are allowed for each player (Pass and Run). This assumption corresponds with the data in that we only observe pass and run for the offense.[19] That said, the assumption does simplify the model substantially while highlighting an important aspect of the situation. In regards to the Defense, it is assumed that there is a "run defense" and a "pass defense". Note that the mixed strategy of the Defense may be better interpreted as the relative strength of the defensive play call in terms of stopping the run or the pass. At the extremes lie the goalline defense for the run and a "dime" package for the pass.[20] Second, the probability for the "wrong" play on the defense is the same irrespective of the actual play call. It would be preferable to allow more flexibility in the model but this assumption allows the success rate on fourth down to be identified. Third, the expected points (*f(.)*, *g(.)*) from each position do not vary with other observable characteristics. This assumption is made for simplicity and for easy comparison with Romer (2006). Fourth, the value of a successful pass is assumed to be 15 yards and the value of a successful run is assumed to be the to go distance. This is an arbitrary assumption but it does correspons do the general view that a pass is a "long play" and a run is a "short play". The choice of 15 yards is made to correspond to the discussion in Romer (2006). Note that changing pass to 20 yards and run to 0 makes little difference to the results. The important

[19] Romer attempts to code the play call rather than what actually happened. One difficulty is the coding quarterback runs. These are assumed to be pass plays.
[20] A dime package includes 6 defensive backs.

difference is the yards gained and it is the change in relative payoffs that may lead to changes going from third down to fourth down.

Given the model, success rates for a given down d are:

Pr(Success | Offense plays Pass, d) = $a\, q_d + b\,(1 - q_d)$

Pr(Success | Offense plays Run, d) = $b\, q_d + c\,(1 - q_d)$

Pr(Success | d) = $a\, p_d\, q_d + b\,(1 - p_d)\, q_d + b\, p_d\,(1 - q_d) + c\,(1 - p_d)(1 - q_d)$

where q_d is the probability the Defense chooses Pass on down d, p_d is the probability the Offense chooses Pass on down d. Note that all the probabilities are functions of X. I have dropped the notation to make the exposition clearer. The probability of success if the Offense plays Pass is the probability of success given both players play Pass (a) times the probability the defense plays Pass (q_d) plus the probability of success given the Offense plays Pass and the Defense plays Run ($1 - q_d$).

In the mixed strategy equilibrium on third down we have

$$q_3\, a\, (f(P+15)-g(P)) + (1-q_3)\, b\, (f(P+15)-g(P)) = q_3\, b\, (f(P+T)-g(P)) + (1-q_3)\, c\, (f(P+T) - g(P))$$

and

$$p_3\, a\, (f(P+15) -g(P)) + (1-p_3)\, b\, (f(P+T)-g(P)) = p_3\, b\, (f(P+15)-g(P)) + (1-p_3)\, c\, (f(P+T) - g(P))$$

In the first case the probability that the Defense plays Pass (q_3) is set such that the Offense is indifferent between playing Pass and Run. In the second case the probability that the Offense plays Pass (p_3) is set such that the Defense is indifferent between Pass and Run. This paper posits that the teams play mixed strategies on third down. It is possible that there are multiple Nash equilibria and this possibility may lead to identification problems (see Berry and Tamer (2005) for example). However, the observation that p_3 is strictly between 0 and 1 suggests a mixed strategy equilibrium is being played. See Chiapoori et al (2002) for tests of mixed strategy equilibrium in zero-sum games in soccer.[21]

If a mixed strategy is assumed on fourth down, we have

$$\begin{aligned} &q_4 a(f(P+15)+f(100-P))+(1-q_4)b(f(P+15)+f(100-P)) \\ &\quad = q_4 b(f(P+T)+f(100-P))+(1-q_4)c(f(P+T)+f(100-P)) \end{aligned}$$

and

$$\begin{aligned} &p_4\, a(f(P+15)+f(100-P))+(1-p_4)b(f(P+T)+f(100-P)) \\ &\quad = p_4 b(f(P+15)+f(100-P))+(1-p_4)\, c(f(P+T)+f(100-P)) \end{aligned}$$

[21] It is less obvious that there is a mixed strategy equilibrium on fourth down. The assumption allows the value of interest to be identified.

Again, in the mixed strategy equilibrium the probabilities that the two players choose Pass (p_4 and q_4) are set such that the other player is indifferent between playing Pass and playing Run.

Given the modeling assumptions can we identify Pr(Success | $d = 4$)? First note that p_3, Pr(Success|Offense plays Pass, $d = 3$), Pr(Success|Offense plays Run, $d = 3$), $f(.)$ and $g(.)$ are observed. Still, given a mixed strategy equilibrium on third down we *cannot* separately identify a, b, c, q_3, and q_4. However, if we assume a mixed strategy equilibrium on fourth down we can identify p_4 and the number of interest Pr(Success | $d = 4$). The appendix presents the algebra solving for this probability as a function of observables.

The paper uses the "two-step" procedure for estimating structural models suggested by Bajari et al (2006). In the first step the paper estimates $p_3(X)$, Pr(Success|Offense plays Pass, $d = 3$, X), Pr(Success|Offense plays Run, $d = 3$, X) separately using probit models for each. In the second step the estimates are put in the equation derived from the structural model to give Pr(Success | $d = 4$, X). The major advantage of the two-step procedure is that it uses very few computer resources while still giving a consistent estimate. The bootstrap is used to calculate the standard errors on Pr(Success | $d = 4$, X).

Chart 5: Probability of Success By To Go Distance at Each Ten Yard Line
(Structural Estimates of 2000 Philadelphia Eagles Home on Week 17
with Third Down Data Q1 1998-2000)

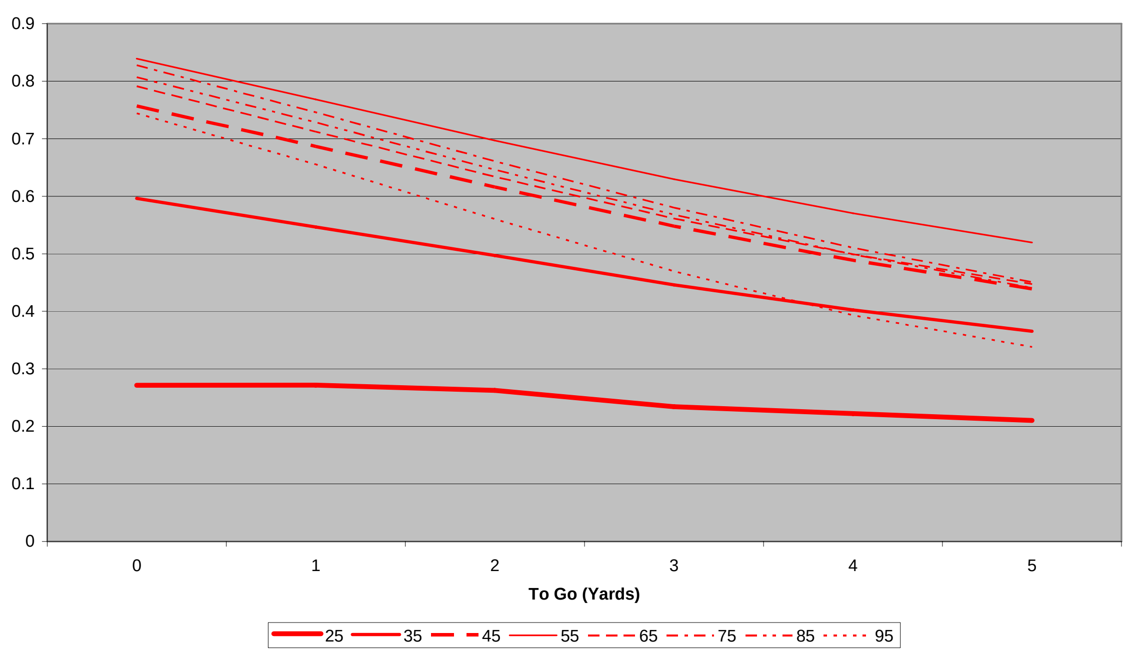

Chart 5 presents the success rates on fourth down for the 2000 Philadelphia Eagles at home on week 17 as a function of the to go distance for each 10 yards on the field starting at the Eagles's own 15 yard line (although positive probabilities only begin at the team's own 25 yard line).[22] The Philadelphia Eagles are used because they have the median third down success rate over all three years. Continuous observable characteristics (spread, over-under) are set at the mean. The to go distances enter into the three probit models quadratically and there is a dummy for each 10 yard distance, 0 to 10, 10 to 20, etc. The three probit models also include dummies for home/away, the year, the week and a fixed effect for each offensive team. See the appendix for the results from the three probit models and the fourth down success rates and the 95% confidence interval.

The results suggest that success rates for going for it in the team's own half of the field are very low (0 below the 20 yard line). For the most part success rates increase as the team gets closer to its own end of the field. The chart suggests there are four distinct areas of the field. Between the team's own 20 and the 50, success rates increase as the team moves closer to their goal. In each case the success falls as the to go distance increases, however the change is not that large although the rate of change is increasing as the field position improves. From the 50 to the opponent's 10 success rates don't change a lot (they may decrease somewhat. Within the opponent's own 10 yard line success rates fall and are lower than for positions further up field.

Determining exactly what is driving these results is difficult because they are based on three different probit models (one for the play call, one for success given pass and one for success given run) all of which flexibly estimate the impact of field position and these results enter into the fourth down success rate in a non-linear way interacted with the expected points from successful passes, successful runs, punts, field goals and turning the ball over on downs. It is not the case that the results simply reflect success rates from the third down plays. Aggregate success rates on third down do not vary significantly across field positions. The results seem to be driven by the interaction between play choice and the success rates conditional on the play call.

Romer (2006) argues that it is fine to use third down data directly because although going from third down to fourth down leads to a change in payoffs the Nash equilibrium

---

[22] The model actually gives small negative numbers for the 15 yard line. These negative numbers are interpreted as 0.

outcome does not change because neither the Offense nor the Defense has a larger influence on the results. The author also states that the Nash equilibrium outcome does not change by much because the payoffs do not change by much. This paper directly calculates how the success rate changes due to the change in payoffs. The change is substantial.

6. Do Teams "Go For It" Enough?

Using the third down data to estimate success rates, Romer (2006) argues that teams should be indifferent between going for it and punting on their own 25 when there are 4 yards or less to go. This calculation is made by comparing the expected number of points from going for it to the expected number of points from punting the ball away. Consider the value of going for it on fourth down, it is

$$\Pr(\text{Success} \mid d = 4)\, f(25) - (1 - \Pr(\text{Success} \mid d = 4)\, f(75)$$

or the probability of making it multiplied by the expected number of points that can be earned from the offenses' own 25 less the probability of turning the ball over times expected number of points the other team will get from the offenses' own 25. Using third down data and the Romer (2006) estimates of $f(.)$ the expected value of going for it with 4 yards to go is approximately .5*1 - .5*3.3 = -1.3. On the other hand if the team punts and has a 40 yard punt the other team must start at their own 35 the expected net points is approximately -1.3. That is, the team should be indifferent between punting and going for it on fourth down when they are at their own 25 with 4 yards to go. Further with 3 yards to go, this analysis suggests teams should *always* go for it. Yet the model presented in the first section using actual data on going for it suggest there is only a 2% probability that a team goes for it on their own 25 with 3 yards to go. Romer (2006) argues that the discrepancy is due to the fact that NFL coaches do not go for it enough. The author provides a number of explanations including ambiguity aversion.

The results from the actual fourth down data and from the simulated fourth down data seem to confirm Romer's result. However, the structural model gives success rates that are substantially lower than suggested by the raw third down data. These success rates are particularly low in the team's own half of the field.

Chart 6: Indifference Curves for Going For It

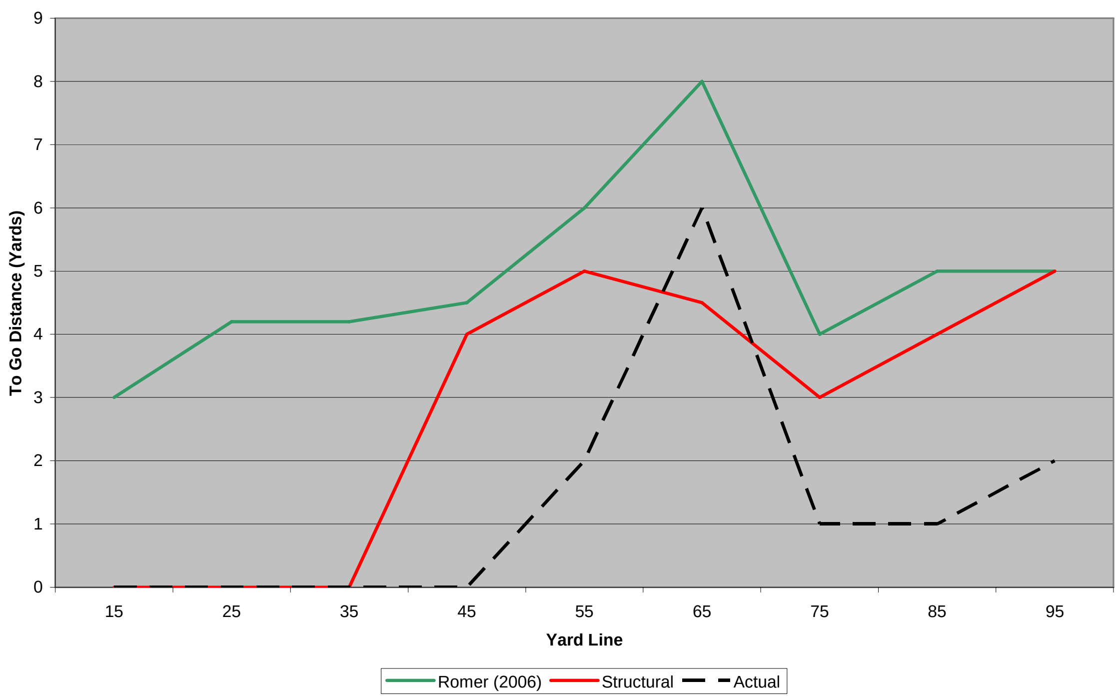

Chart 6 presents the "indifference curves" for an NFL coach. At a given field position and to go distance any place below the line means the team should go for it and any place above the line means the team should kick. The Romer (2006) and Actual lines come from Figure 5 in Romer (2006). The Romer (2006) line is calculated using success rates from the raw third down data. The Actual line is the point at which actual teams are equally likely to go for it as they are punt from the actual fourth down data. The Structural line is calculated as the point at which the expected points from going for it is equal to the expected points from kicking using success rates from the structural model. These lines are calculated at the 5, 15, 25, etc.

The chart highlights Romer's point that teams do not go for it enough on fourth down. In particular, Romer (2006) suggests that a team on its own 10 yard line should be indifferent between going for it and punting when there is 3 yards to go for a first down. From the actual data we see that teams only have equal probabilities of going for it and punting in their own half of the field. The structural model presented above suggests that

Romer (2006) overstates the advantage of going for it on fourth down, particularly in the team's own half of the field. The structural model predicts teams should go for it more than they do around the half and in the Red Zone. The structural model predicts teams should go for it less than they do when the ball is between the 40 and the 30.

Conclusion

Romer (2006) argues that NFL coaches do not go for it enough on fourth down and suggests this is evidence that some economic actors do not behave as rationally as many economic models assume. One difficult with determining the value of going for it on fourth down is that it is necessary to know the probability of success. The difficulty is that we rarely observe actual fourth down attempts in the NFL. To overcome the lack of data Romer (2006) uses actual success rates from third downs. This paper presents three alternative strategies for overcoming the lack of data. The first is to use actual fourth down attempts and various assumptions to account for missing or biased data. The second strategy is to use data from simulated fourth down attempts. The third strategy is to use the third down data to estimate a structural model and calculate equilibrium success rates on fourth down. The results from these three strategies are then compared with using third down success rates. The results from the actual fourth down data and the simulated data are similar to those derived from the third down data. The results from the structural model are substantially different from the third down results particularly for success rates in the team's own half of the field. The results from the structural model are also more consistent with the observed behavior of actual NFL teams suggesting that NFL coaches may in fact be or act as if they are rational dynamic optimizers. Further, firms may actually maximize profits.

Reference

Adams, Christopher, 2006, "Learning in Prediction Markets," SSRN Working Paper at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=923155.

Bajari, Patrick, Lanier Benkard and Jon Levine, 2006, "Estimating Dynamic Models of Imperfect Competition," Working Paper at http://www.econ.umn.edu/~bajari/research/published_papers/dynamic[1].pdf.

Berry, Steve, James Levinsohn and Ariel Pakes, 1995, "Automobile Prices in Equilibrium," *Econometrica*, 60(4): 889-917.

Berry, Steve and Elie Tamer, "Identification in Models of Oligopoly Entry," Working paper at http://www.econ.yale.edu/~steveb/wc.pdf.

Camerer, Colin, 2003, *Behavioral Game Theory: Experiments in strategic interaction*, Princeton University Press.

Chiappori, Pierre-Andre, Timothy Groseclose and Steve Levitt, 2002, "Testing Mixed-Strategy Equilibria When Players Are Heterogeneous: The Case of Penalty Kicks in Soccer." *American Economic Review*, *92*, pp. 1138-1151.

Greene, William, 2003, *Econometric Analysis*, Prentice Hall, 5th Edition.

Levitt, Steve, 2006, "An Economist Sells Bagels: A case study in profit maximization," *NBER Working Paper W12152*.

Manski, Charles, 1995, *Identification Problems in the Social Sciences*, Harvard University Press.

Romer, David, 2006, "Do Firms Maximize? Evidence from professional football," *Journal of Political Economy*, April, at http://www.econ.berkeley.edu/users/dromer/papers/PAPER_NFL_JULY05_FORWEB_CORRECTED.pdf (page numbers are cited from this version of the paper).

Rust, John, 1986, "Structural Estimation of Markov Decision Processes," *Handbook of Econometrics*, Eleviser, 4: 3081-3143.

Appendix 1.

Table A1: Probability of Success (Selection Adjusted Probit - Heckit) (All Fourth Downs 1998-2000)

	Coef.	Std. Err.
Probit on Success		
To Go Distance (Yards)	-0.11	0.02
To Go Squared (10xYards)	0.03	0.01
Constant	0.39	0.08
Selection Probit		
Less Than 1.5 Yards (Dummy)	1.32	0.06
To Go Distance (Yards)	-0.04	0.01
Yard Line 0 to 10 (Dummy)	-1.04	0.27
Yard Line 10 to 20 (Dummy)	-0.89	0.13
Yard Line 20 to 30 (Dummy)	-1.10	0.11
Yard Line 30 to 40 (Dummy)	-0.80	0.10
Yard Line 40 to 50 (Dummy)	-0.58	0.09
Yard Line 50 to 60 (Dummy)	0.07	0.08
Yard Line 60 to 70 (Dummy)	0.93	0.08
Yard Line 70 to 80 (Dummy)	0.30	0.08
Yard Line 80 to 90 (Dummy)	-0.22	0.10
Points Up (Points)	-0.03	0.00
Points Up Late (Points)	-0.02	0.00
Minutes Remaining in Quarter (Minutes)	-0.02	0.01
Minutes Remaining in Quarter Late (Minutes)	-0.12	0.01
First Quarter (Dummy)	-2.13	0.10
Second Quarter (Dummy)	-1.84	0.08
Third Quarter (Dummy)	-1.76	0.09
Point Spread (Points)	0.00	0.01
Over-under (Points)	0.02	0.01
Home (Dummy)	0.03	0.05
Week (Integer 1 to 17)	0.02	0.00
1998 (Dummy)	0.08	0.05
1999 (Dummy)	-0.04	0.05
Offense and Defense Dummies	Yes	
Constant	-0.88	0.38
rho	-0.04	0.06
N	11,358	
N (Uncensored)	1,332	
Log Likelihood	-3,134.55	

Table A2: Probit Regressions on Third Down Data (1998-2000)

	Pass		Success Given Pass		Success Given Run	
	Coefficient	SE	Coefficient	SE	Coefficient	SE
To Go (Yards)	0.33	0.01	-0.09	0.01	-0.22	0.03
To Go Squared (Yards)	-0.01	0.00	0.00	0.00	0.01	0.00
10 Yard Line or Less (Dummy)	-0.62	0.19	0.25	0.18	-0.52	0.50
10 to 20 Yard Line (Dummy)	0.10	0.15	0.39	0.13	0.08	0.31
20 to 30 Yard Line (Dummy)	0.22	0.12	0.32	0.11	-0.05	0.22
30 to 40 Yard Line (Dummy)	0.34	0.11	0.39	0.11	-0.15	0.22
40 to 50 Yard Line (Dummy)	0.29	0.11	0.38	0.11	0.08	0.21
50 to 60 Yard Line (Dummy)	0.21	0.12	0.52	0.11	0.05	0.22
60 to 70 Yard Line (Dummy)	0.30	0.13	0.34	0.12	0.02	0.26
70 to 80 Yard Line (Dummy)	-0.02	0.12	0.31	0.12	0.12	0.23
80 to 90 Yard Line (Dummy)	-0.06	0.14	0.32	0.13	0.15	0.27
Points Spread (Points)	-0.01	0.01	-0.01	0.00	0.00	0.01
Over-under (Points)	0.01	0.01	0.00	0.01	0.01	0.02
Home (Dummy)	0.11	0.06	-0.12	0.05	0.21	0.12
1998 (Dummy)	0.01	0.07	0.05	0.05	-0.18	0.14
1999 (Dummy)	-0.05	0.07	0.01	0.05	0.05	0.14
Week Dummies	Yes		Yes		Yes	
Offense Dummies	Yes		Yes		Yes	
Constant	-0.99	0.37	-0.02	0.29	0.92	0.77
Log Likelihood	-1,376.46		-2,591.71		-371.82	
Pseudo R Squared	0.27		0.06		0.17	
N	4,729		4,080		649	

Table A3: Bootstrapped Confidence Intervals for Chart 5

	25		35		45		55	
Togo	2.5%	97.5%	2.5%	97.5%	2.5%	97.5%	2.5%	97.5%
0	0.21	0.30	0.46	0.66	0.61	0.84	0.68	0.92
1	0.21	0.31	0.43	0.61	0.55	0.78	0.61	0.86
2	0.21	0.30	0.40	0.57	0.49	0.71	0.57	0.80
3	0.18	0.28	0.36	0.53	0.42	0.65	0.51	0.74
4	0.18	0.27	0.33	0.50	0.38	0.60	0.46	0.69
5	0.17	0.27	0.30	0.47	0.35	0.57	0.41	0.64

	65		75		85		95	
Togo	2.5%	97.5%	2.5%	97.5%	2.5%	97.5%	2.5%	97.5%
0	0.62	0.88	0.64	0.90	0.62	0.89	0.55	0.85
1	0.56	0.81	0.58	0.82	0.55	0.82	0.47	0.76
2	0.49	0.73	0.51	0.74	0.48	0.74	0.39	0.65
3	0.44	0.67	0.45	0.67	0.41	0.67	0.33	0.55
4	0.39	0.62	0.40	0.62	0.35	0.61	0.27	0.48
5	0.34	0.58	0.36	0.58	0.31	0.56	0.22	0.44

Notes:

1. All negatives are set to 0.

2. Some of the bootstrapped samples the success rates are not calculated because Stata drops a variable in one of the probit regressions (162 in total). The percentiles are taken at the closest integer to 2.5% and 97.5% respectively.

Appendix 2. Structural Model of Third Down

Consider the following zero-sum game. The Offense and the Defense can simultaneously choose between two actions, Run and Pass. The probabilities of success (for the Offense) from each combination of actions are denoted:

Pr(Success | Offense plays Pass, Defense plays Pass) = a

Pr(Success | Offense plays Pass, Defense plays Run) = b

Pr(Success | Offense plays Run, Defense plays Pass) = b

Pr(Success | Offense plays Run, Defense plays Run) = c

The third down payoff matrix is:

		Defense	
		Pass	*Run*
Offense	*Pass*	$a(f(P+15) - g(P))$	$b(f(P+15) - g(P))$
	Run	$b(f(P+T) - g(P))$	$c(f(P+T) - g(P))$

where $f(P)$ is the expected number of points from first down at a particular position P, $g(P)$ is the expected number of points from kicking (punting or kicking a field goal) and T is the to go distance. As this is a zero-sum game the payoffs are those to the Offense, while the Defense payoffs are simply the negative.

The fourth down payoff matrix:

		Defense	
		Pass	*Run*
Offense	*Pass*	$a(f(P+15) + f(100 - P))$	$b(f(P+15) + f(100 - P))$
	Run	$b(f(P+T) + f(100 - P))$	$c(f(P+T) + f(100 - P))$

Given the model, success rates are:

(1) Pr(Success | Offense plays Pass, $d = 3$) = $S_p = a\, q_3 + b\,(1 - q_3)$

(2) Pr(Success | Offense plays Run, $d = 3$) = $S_r = b\, q_3 + c\,(1 - q_3)$

(3) Pr(Success | $d = 4$) = $a\, p_4\, q_4 + b\,(1 - p_4)\, q_4 + b\, p_4\,(1 - q_4) + c\,(1 - p_4)(1 - q_4)$

where q_d is the probability the Defense chooses Pass on down d, p_d is the probability the Offense chooses Pass on down d.

In the mixed strategy equilibrium on third down we have

(4) $q_3\, a\, (f(P+15) - g(P)) + (1-q_3)\, b\, (f(P+15) - g(P)) = q_3\, b\, (f(P+T) - g(P)) + (1-q_3)\, c\, (f(P+T) - g(P))$

and

(5) $p_3\, a\, (f(P+15) - g(P)) + (1-p_3)\, b\, (f(P+T) - g(P)) = p_3\, b\, (f(P+15) - g(P)) + (1-p_3)\, c\, (f(P+T) - g(P))$

In a mixed strategy is assumed on fourth down, we have the following equalities.

(6) $q_4 a(f(P+15)+f(100-P))+(1-q_4)b(f(P+15)+f(100-P))$
$=q_4 b(f(P+T)+f(100-P))+(1-q_4)c(f(P+T)+f(100-P))$

and

(7) $p_4\, a(f(P+15)+f(100-P))+(1-p_4)\, b(f(P+T)+f(100-P))$
$=p_4\, b(f(P+15)+f(100-P))+(1-p_4)\, c(f(P+T)+f(100-P))$

Rearranging (1)

(1') $S_p - b = (a - b)\, q_3$

Rearranging (2)

(2') $c - S_r = (c - b)\, q_3$

Dividing (1') by (2') we have

(8) $(S_p - b)/(c - S_r) = (a - b)/(c - b)$

Let $A = f(P+15)$, $B = f(P+T)$, $C = g(P)$. Rearranging (5) we have

(5') $(a - b)/(c - b) = ((1-p_3)/p_3)\, ((B - C)/(A - C))$

Substituting (8) into (5') we have the relationship between c and b,

(9) $c = S_r + (S_p - b)\, (p_3 / (1-p_3))\, ((A - C)/(B - C))$

Rearranging (5')

(5'') $(a - b) = c((1-p_3)/p_3)\, ((B - C)/(A - C)) - b((1-p_3)/p_3)\, ((B - C)/(A - C))$

Substituting in (9) we have the relationship between a and b

(10) $a = S_p + (S_r - b)((1-p_3)/p_3)\, ((B - C)/(A - C))$

Letting D = $-f(100-P)$ and rearranging (7)

(7') $p_4 = 1/(((a - b) / (c - b))((A - D) / (B - D)) + 1)$

Substituting in (5') we have p_4 as a function of observables

(11) $p_4 = 1/(((1-p_3)/p_3)\, ((B - C)/(A - C))((A - D) / (B - D)) + 1)$

Rearranging (6) we have

(6') $q_4 = (c(B - D) - b(A - D))/(a(A - D) - b(A - D) - b(B - D) + c(B - D))$

Let S_4 = Pr(Success | d = 4) and rearranging (3) we have

(3') $S_4 = (a - 2b + c)\, p_4\, q_4 + (b - c)\, q_4 + (b - c)\, p_4 + c$

Substituting (9), (10), (11) and (6') into (3') to get S_4 as a function of observables

(12) $$S_4 = ((CD^2 - A^2(B - C)(1 - p_3) + B^2 C p_3 + B D(-D(1 - p_3) - 2Cp_3) - A(-2BD + B^2 p_3 + D(2C - 2Cp_3 + Dp_3)))(-Ap_3 S_p - B(1 - p_3)S_r + C(p_3 S_p + S_r - p_3 S_r)))/(CD + B(-D(1 - p_3) - Cp_3) + A(B - C(1 - p_3) - Dp_3))^2$$

It is not clear exactly why S_4 is identified in this model given we have 3 equations and 4 unknowns but the algebra works out.