DO FIRMS MAXIMIZE? EVIDENCE FROM PROFESSIONAL FOOTBALL

David Romer

University of California, Berkeley

Revised, July 2005

ABSTRACT

This paper examines a single, narrow decision—the choice on fourth down in the National Football League between kicking and trying for a first down—as a case study of the standard view that competition in the goods, capital, and labor markets leads firms to make maximizing choices. Play-by-play data and dynamic programming are used to estimate the average payoffs to kicking and trying for a first down under different circumstances. Examination of teams' actual decisions shows systematic, clear-cut, and overwhelmingly statistically significant departures from the decisions that would maximize teams' chances of winning. Possible reasons for the departures are considered.

I am indebted to Ben Allen, Laurel Beck, Sungmun Choi, Ryan Edwards, Mario Lopez, Peter Mandel, Travis Reynolds, Evan Rose, and Raymond Son for outstanding research assistance, to Christina Romer for invaluable discussions, and to numerous colleagues and correspondents for helpful comments and suggestions. An earlier version of the paper was titled "It's Fourth Down and What Does the Bellman Equation Say? A Dynamic-Programming Analysis of Football Strategy."

I. INTRODUCTION

A central assumption of most economic models is that agents maximize simple objective functions: consumers maximize expected utility, and firms maximize expected profits. The argument for this assumption is not that it leads to perfect descriptions of behavior, but that it leads to reasonably good approximations in most cases.

The assumption that consumers successfully maximize simple objective functions frequently makes predictions about how a single individual will behave when confronted with a specific, easily describable decision. Thus it can often be tested in both the laboratory and the field. As a result, there is a growing body of evidence about when the assumption does or does not yield reasonably accurate descriptions of behavior.

The assumption that firms maximize profits is much more difficult to test. Particularly for large firms, the decisions are usually complicated and the data difficult to obtain. The a priori case for firm maximization is much stronger than that for consumer maximization, however. As Alchian (1950), Friedman (1953), Becker (1957), Fama (1980), and others explain, competition in the goods, capital, and labor markets creates strong forces driving firms toward profit-maximization. A firm that fails to maximize profits is likely to be outcompeted by more efficient rivals or purchased by individuals who can obtain greater value from it by pursuing profit-maximizing strategies. And managers who fail to maximize profits for the owners of their firms are likely to be fired and replaced by ones who do. Thus the case for firm maximization rests much more on logical argument than empirical evidence. As Friedman puts it, "unless the behavior of businessmen in some way or other approximated behavior consistent with the maximization of returns, it seems unlikely that they would remain in business for long. … The process of 'natural selection' thus helps to validate the hypothesis [of return-maximization]" (1953, p. 22).

This paper takes a first step toward testing the assumption that firms maximize profits. It does so by looking at a specific strategic decision in professional sports: the choice in football between kicking and trying for a first down on fourth down. Examining strategic decisions in sports has two enormous advantages. First, in most cases, it is difficult to think of any significant channel through which strategic decisions are likely to affect a team's profits other than through their impact on the team's probability of winning. Thus, the problem of maximizing profits plausibly reduces to the much simpler problem of maximizing the probability of winning. Second, there are copious, detailed data describing the circumstances teams face when they make these decisions.[1]

The predictions of simple models of optimization appear especially likely to hold in the case of fourth-down decisions in professional football. There are three reasons. First, the market for the coaches who make these decisions is intensively competitive. Salaries average roughly $3 million per year, and annual turnover exceeds 20 percent.[2] Second, winning is valued enormously (as shown by the very high salaries commanded by high-quality players). And third, the decisions are unusually amenable to learning and imitation: the decisions arise repeatedly, and information about others' decisions is readily available. Thus a failure of maximization in this setting would be particularly striking.

This paper shows, however, that teams' choices on fourth downs differ from the choices that would maximize their chances of winning in ways that are systematic and overwhelmingly statistically significant. Indeed, there are cases where teams consistently make choices that represent clear-cut and large departures from win-maximization.

One case where the departure from win-maximizing choices is particularly striking and relatively easy to see arises when a team faces fourth down and goal to go on its opponent's 2-yard

---

[1] Thaler (2000) stresses the potential value of sports decision-making in testing the hypothesis of firm optimization.

[2] The salary figure is based on the 23 coaches (out of 32) for whom 2004 salary information could be obtained from publicly available sources. The turnover data are for 1998-2004.

line early in the game.[3] In this situation, attempting a field goal is virtually certain to produce three points, while trying for a touchdown has about a three-sevenths chance of producing seven. Thus the two choices have essentially the same expected payoff in terms of immediate points. But if the team tries for a touchdown and fails, its opponent typically gains possession of the ball on the two-yard line; if the team scores a touchdown or makes a field goal, on the other hand, the opponent returns a kickoff, which is on average considerably better for it. Thus trying for a touchdown on average leaves the opponent in considerably worse field position. I show later that rational risk aversion about points scored, concern about momentum, and other complications do not noticeably affect the case for trying for a touchdown. As a result, my estimates imply that the team should be indifferent between attempting a field goal and trying for a touchdown if the probability of scoring a touchdown is about 18 percent. They also imply that trying for a touchdown rather than attempting a field goal would increase the team's chances of winning the game by about 3 percentage points, which is very large for a single play. In fact, however, teams attempted a field goal all 9 times in my sample they were in this position.

Analyzing the choice between kicking and trying for a first down or touchdown in other cases is more complicated: the immediate payoffs in terms of expected scoring may be different under the two choices, and the attractiveness of the distributions of ball possession and field position may be difficult to compare. Fortunately, however, the problem can be analyzed using dynamic programming. The choice between kicking and going for it leads both to an immediate payoff in terms of points (which may be zero) and to one team having a first down somewhere on the field. That first down leads to additional scoring (which again may be zero) and to another possession and first down. And so on.

Section II of the paper therefore uses data from over 700 National Football League games to estimate the values of having a first down at any point on the field (as well as the value of kicking off). To avoid the complications introduced when one team is well ahead or when the end

[3] The appendix summarizes the rules of football that are relevant to the paper.

of a half is approaching, I focus on the first quarter.

Section III uses the results of this dynamic-programming analysis to examine teams' decisions on fourth down over the entire field. To estimate the value of kicking at various points on the field, I use the outcomes of actual field-goal attempts and punts. Decisions to go for it on fourth down (that is, not to kick) are sufficiently rare, however, that they cannot be used to estimate the value of trying for a first down or touchdown. I therefore use the outcomes of third-down plays instead. I then compare the values of kicking and going for it to determine which decision is better on average as a function of where the team is on the field and the number of yards it needs for a first down or touchdown. Finally, I compare the results of this analysis with teams' actual choices. I find that over the entire field, teams' choices are far more conservative than the ones that would maximize their chances of winning.

Section IV considers possible complications and biases. I examine the possibility of rational risk aversion, sample-selection bias in the estimation of the values of different situations, possible differences between third-down plays and fourth-down attempts, the possession of additional information by teams, concern about momentum, and general-equilibrium effects. I find that none of these complications change the basic messages of the analysis.

Section V considers the results' quantitative implications. Because the analysis concerns only a small fraction of plays, it implies that different choices on those plays could have only a modest impact on a team's chances of winning. But it also implies that there are circumstances where teams essentially always kick even though the case for going for it is clear-cut and the benefits of going for it are substantial.

Finally, Section VI discusses the results' broader implications. As I describe, the departures from win-maximization could reflect either a pursuit of a different, more complex objective function or a failure of maximization. I discuss how either of these possibilities might arise and how one might be able to distinguish between them.[4]

[4] Many papers use sports data to test economic theories. Attempts to use economic tools to analyze sports strategy—and in doing so to use sports data to test hypotheses about maximization—are less common. Two recent examples are Walker and Wooders (2001), who

II. THE VALUES OF DIFFERENT SITUATIONS

Framework. The dynamic-programming analysis focuses on 101 situations: a first down and 10 on each yard line from a team's 1 to its opponent's 10, a first and goal on each yard line from the opponent's 9 to its 1, a kickoff from the team's 30 (following a field goal or touchdown, or at the beginning of the game), and a kickoff from its 20 (following a safety). Let V_i denote the value of situation i. Specifically, V_i is the expected long-run value, beginning in situation i, of the difference between the points scored by the team with the ball and its opponent when the two teams are evenly matched, average NFL teams.

By describing the values of situations in terms of expected point differences, I am implicitly assuming that teams are risk-neutral over points scored. This is clearly not a good approximation late in a game: a team trailing by two points with time running out is not indifferent between three points for sure and a three-sevenths chance of seven. But as I show in Section IV, it is an excellent approximation for the early part of the game. For that reason, I focus on the first quarter.

Focusing on the first quarter has a second advantage: it makes it reasonable to neglect effects involving the end of a half. Since play stops at the end of each half, the value of a first down on one's 20-yard line with a minute left in a half may be quite different from the value of the same situation with three minutes left. But because play in the second quarter begins at the

---

examine serves in tennis, and Chiappori, Levitt, and Groseclose (2002), who consider penalty kicks in soccer. In contrast to this paper, these papers find no evidence of large departures from optimal strategies.

Three studies are more closely related to this paper. Carter and Machol (1971) propose and implement a recursive approach to estimating the value of having the ball at different points on the field. Carter and Machol (1978) and Carroll, Palmer, and Thorn (1998, Ch. 10) examine fourth-down decisions systematically. In broad terms, many of their steps are similar to mine. The specifics of their analyses, however, are quite different and considerably cruder. The same is true of their findings: both studies also conclude that teams' fourth-down decisions are considerably more conservative than the decisions that would maximize their chances of winning, but their exact conclusions differ substantially from mine. I discuss the particulars of how my analysis is related to these studies below.

point where the first quarter ended, the value of a first down on one's 20 with a minute left in the quarter is almost certainly very close to the value of the same situation with three minutes.

Let g index games and t index situations within a game. Let D^i_{gt} be a dummy variable that equals 1 if the t^{th} situation in game g is a situation of type i. For example, suppose that i = 100 denotes a kickoff from one's 30; then, since all games begin with a kickoff, D^{100}_{g1} = 1 for all g and D^i_{g1} = 0 for all g and for all i ≠ 100. Let P_{gt} denote the net points scored by the team with the ball in situation g,t before the next situation. That is, P_{gt} is the number of points scored by the team with the ball minus the number scored by its opponent. Finally, let B_{gt} be a dummy that equals 1 if the team with the ball in situation g,t also has the ball in situation g,t+1 and that equals -1 if the other team has the ball in situation g,t+1.

Now consider the team with the ball in situation g,t. The situation's realized value to the team as of one situation later has two components. The first is the net points it scores before the next situation; this is P_{gt}. The second is the value of the new situation. If the team has the ball in that situation, the value to it of the new situation is simply the V_i corresponding to that situation. If the other team has the ball, the value of the new situation to the team that had the ball in situation g,t is minus the V_i corresponding to the new situation (since the value of the new situation to the team without the ball is equal and opposite to the value of the situation to its opponent). In terms of the notation just introduced, the value of situation g,t+1 to the team with the ball in situation g,t equals the one of the 101 D^i's that equals one in that situation, times the V_i for that situation, times B_{gt}. That is, it equals $B_{gt}\sum_i D^i_{gt+1} V_i$.

The value of situation g,t as of that situation must equal the expectation of the situation's realized value one situation later. We can write the value of situation g,t conditional on being in that situation as $\sum_i D^i_{gt}$. Thus we have

$$\sum_i D^i_{gt} V_i = E[P_{gt} + B_{gt}\sum_i D^i_{gt+1} V_i], \qquad (1)$$

where the expectation is conditional on situation g,t.

Now define e_{gt} as the difference between the realized value of situation g,t one situation later and the expectation of the realized value conditional on being in situation g,t: $e_{gt} = [P_{gt} + B_{gt} \sum_i D^i_{gt} V_i] - E[P_{gt} + B_{gt} \sum_i D^i_{gt+1} V_i]$. By construction, e_{gt} is uncorrelated with each of the D^i_{gt}'s. If e were correlated with a D^i, this would mean that when teams were in situation i, the realized value one situation later would differ systematically from V_i; but this would contradict the definition of V_i.

Using this definition of e_{gt}, we can rewrite (1) as

$$\sum_i D^i_{gt} V_i = P_{gt} + B_{gt} \sum_i D^i_{gt+1} V_i - e_{gt}, \tag{2}$$

or

$$\begin{aligned} P_{gt} &= \sum_i D^i_{gt} V_i - B_{gt} \sum_i D^i_{gt+1} V_i + e_{gt} \\ &= \sum_i V_i (D^i_{gt} - B_{gt} D^i_{gt+1}) + e_{gt}. \end{aligned} \tag{3}$$

To think about estimating the V_i's, define $X^i_{gt} = D^i_{gt} - B_{gt} D^i_{gt+1}$. Then (3) becomes

$$P_{gt} = \sum_i V_i X^i_{gt} + e_{gt}. \tag{4}$$

This formulation suggests that possibility of estimating the V_i's by regressing P on the X's. But e may be correlated with the X's. Specifically, e_{gt} is likely to be correlated with the $-B_{gt} D^i_{gt+1}$ terms of the X^i_{gt}'s. Recall, however, that e_{gt} is uncorrelated with the D^i_{gt}'s. Thus the D^i_{gt}'s are legitimate instruments for the X^i_{gt}'s. Further, since they enter into the X^i_{gt}'s, they are almost surely correlated with them. Thus we can estimate (4) by instrumental variables, using the D^i_{gt}'s as the instruments.[5]

---

[5] There is another way of describing the estimation of the V_i's. Begin with an initial set of V_i's

There is one final issue. There are 101 V_i's to estimate. Even with a large amount of data, the estimates of the individual V_i's will be noisy. But the value of a first down is almost certainly a smooth function of a team's position on the field. If this is correct, forcing the estimates of the V_i's to be smooth will improve the precision of the estimates while introducing minimal bias. I therefore require the estimated V_i's for first downs to be a quadratic spline as a function of the team's position on the field, with knot points at both 9, 17, and 33 yard lines and at the 50. I do not impose any restrictions on the two estimated V_i's for kickoffs. The imposition of the spline reduces the effective number of parameters to be estimated from 101 to 12.[6]

Data and Results. Play-by-play accounts of virtually all regular season National Football League games for the 1998, 1999, and 2000 seasons were downloaded from the NFL website, nfl.com.[7] Since I focus on strategy in the first quarter, I only use data from first quarters to estimate the V_i's.

The 732 regular season games for which play-by-play accounts are available yield a total of 11,112 first-quarter situations. By far the most common situations are a kickoff from one's 30-yard line (1851 cases) and a first and 10 on one's 20 (557 cases). Because 98.4 percent of extra-point attempts were successful in this period, all touchdowns are counted as 6.984 points.[8]

(such as $V_i = 0$ for all i). Now for each i, compute the mean of the realized values of all situations of type i one situation later using the assumed V_i's and the actual P_{gt}'s. Repeat the process using the new V_i's as an input, and iterate until the process converges. One can show that this procedure produces results that are numerically identical to those of the instrumental-variables approach.

[6] Carter and Machol (1971) also use a recursive approach to estimate point values of first downs at different positions on the field, using a considerably smaller sample from 1969. There are two main differences between their approach and mine. First, they truncate the value associated with a given field position at the time of the first subsequent score; that is, they arbitrarily assign a value of zero to kickoffs and free kicks. Second, they divide the field into 10-yard intervals and estimate the average value for each interval.

[7] Data for two games in 1999 and two games in 2000 were missing from the website.

[8] Note that I am estimating the V_i's, the expected values of situations for an average team, using the average values of situations across observations from all teams. Although the latter is conceptually different from the former, I present evidence in Section IV that it is an excellent approximation.

Figure 1 reports the results of the instrumental-variables estimation. It plots the estimated V for a first and 10 (or first and goal) as a function of the team's position on the field, together with the two-standard-error bands.

The estimated value of a first and 10 on one's 1-yard line is -1.6 points. V rises fairly steeply from the 1, reaching 0 at about the 15. That is, the estimates imply that a team should be indifferent between a first and 10 on its 15 and having its opponent in the same situation. V increases approximately linearly after the 15, rising a point roughly every 18 yards. The value of a first and 10 equals the value of receiving a kickoff from the 30—0.6 points—around the 27-yard line. That is, receiving a kickoff is on average as valuable as a first and 10 on one's 27.

Since a kickoff has a value of -0.6 points, the net value of a field goal is 2.4 points, and the net value of a touchdown is 6.4 points. The value of a first and 10 reaches the net value of a field goal at the opponent's 41-yard line. Finally, V begins to increase more rapidly around the opponent's 10. The estimated value of a first and goal on the 1 is 5.55 points; this is about the same as the value of an 80 percent chance of a touchdown and a 20 percent chance of a field goal. The V's are estimated relatively precisely: except in the vicinity of the goal lines, their standard errors are less than 0.1.

Figure 2 reports the results when the V's are not constrained to be smooth. For comparison, the figure also shows the point estimates for the constrained case. As one would expect, the unconstrained estimates of the V's are much more variable than the constrained ones. They are also much less plausible; for example, they imply that the value of a first down often falls as a team moves closer to its opponent's goal line. Aside from the noisiness of the unconstrained estimates, however, the unconstrained and constrained estimates do not differ in any evident systematic way.

III. KICKING VERSUS GOING FOR IT

This section uses the results of Section II to analyze the choice between kicking and going for it on fourth down. The analysis proceeds in four steps. The first two estimate the values of kicking and going for it in different circumstances. The third compares the two choices to determine which is on average better as a function of the team's position on the field and its distance from a first down. The final step compares teams' actual decisions with the choices that the analysis suggests are preferable.

Kicking. My method for analyzing the values of kicks is similar to the approach in the previous section. I focus on the realized values of kicks as of the subsequent situation (where "situation" is defined as before). This realized value has two components, the net points scored before the next situation and the next situation's value.

If we neglect the issue of smoothing the estimates, the analysis is straightforward. To estimate the value of a kick from a particular yard line, one simply averages the realized values of the kicks from that yard line as of the subsequent situation. In contrast to the previous section, there is no need for instrumental-variables estimation.

One can describe the procedure formally using notation like the previous section's. Let K_i denote the value of a kick from yard line i, and let t index kicks within a game. In addition, let A^i_{gt} be a dummy that equals 1 if kick g,t is from yard line i, $\tilde{D}^i_{gt}$ a dummy that equals 1 if the next situation after kick g,t is a situation of type i, $\tilde{P}_{gt}$ the net number of points scored by the kicking team between kick g,t and the subsequent situation (including any points scored on the kick itself), and $\tilde{B}_{gt}$ a dummy that equals 1 if the kicking team has the ball in the next situation and -1 if the other team has the ball. Proceeding along the lines used to derive equation (3) yields

$$\tilde{P}_{gt} + \tilde{B}_{gt} \sum_i \tilde{D}^i_{gt} V_i = \sum_i K_i A^i_{gt} + u_{gt} . \qquad (4)$$

Here u_{gt} is the difference between the realized value of kick g,t as of the subsequent situation and the expectation of that value conditional on the position the team is kicking from. This definition of u_{gt} implies that it is uncorrelated with the A^i_{gt}'s. Thus (5) can be estimated by OLS; this is equivalent to the averaging procedure just described.

I constrain the estimated values of kicks to be smooth in the same way as before. That is, I require the K_i's to be a quadratic spline as a function of the team's position on the field, with the same knot points as in Section II. I make one modification, however. Teams' choices between punting and attempting a field goal change rapidly around their opponents' 35-yard line. Since one would expect the level but not the slope of the value of kicking as a function of the yard line to be continuous where teams switch from punts to field-goal attempts, I do not impose the slope restriction at the opponent's 33. And indeed, the estimates reveal a substantial kink at this knot point.

The data consist of all kicks in the first quarters of games. Since what we need to know is the value of deciding to kick, I include not just actual punts and field-goal attempts, but blocked and muffed kicks and kicks nullified by penalties. There are 2560 observations.[9]

The results are reported in Figure 3. Panel (a) shows the estimated value of kicking as a function of the team's position on the field, together with the two-standard-error bands.[10] On the

---

[9] There are several minor issues involving the data. First, fourth-down plays that are blown dead before the snap and where the play-by-play account does not say whether the kicking squad was sent in are excluded on the grounds that it is not possible to determine whether the team intended to kick. Since such plays are also excluded from the analysis of the decision to go for it, this exclusion should generate little bias. Second, if a penalty causes one fourth-down play to immediately follow another, both are included. Third, it is not clear whether fake punts and field-goal attempts should be included; it depends on whether one wants to estimate the value of deciding to kick or the value of lining up to kick. There are only five fake kicks in the sample, however, and the results are virtually unaffected by whether they are included. The results reported in the text include fakes. Finally, since teams occasionally obtain first downs on kicking plays (primarily through penalties), the value of a kick is affected by the number of yards the team has to go for a first down. But there are only six kicking plays in the sample where the team had five or fewer yards to go and moved the ball five or fewer yards and obtained a first down. Thus to improve the precision of the estimates, I do not let the estimated value of kicks vary with the number of yards needed for a first down.

[10] The standard errors account for the fact that the V_i's in (5) are estimated, This calculation is

team's 1-yard line, the estimated value of kicking is -2.8; this is the same as the value of the other team having a first and 10 on one's 34. The value rises steadily, and reaches 2.4 at the opponent's 1; this is essentially identical to the value of making a field goal for sure.

Panel (b) presents the results in a way that may be more useful. It plots the difference between the estimated values of a kick and of the other team having a first down on the spot. From the team's 10-yard line to midfield, this difference is fairly steady at around 2.1 points, which corresponds to a punt of about 38 yards. The difference dips down in the "dead zone" around the opponent's 35-yard line, where a field-goal attempt is unlikely to succeed and a punt is likely to produce little yardage. It reaches a low of 1.5 (a punt of only 25 yards) at the opponent's 33, and then rises to 2.2 at the opponent's 21. As the team gets closer to the goal line, the probability of a successful field goal rises little while the value of leaving the opponent with the ball rises considerably. The difference between the values of kicking and of the opponent receiving the ball therefore falls, reaching 0.7 at the 1. The estimates are relatively precise: the standard error of the difference in values is typically about 0.1.

Going for It. The analysis of the value of not kicking on fourth down parallels the analysis of kicking. There are two differences. First, because teams rarely go for a first down or touchdown on fourth down, I use third-down plays to estimate the value of going for it. That is, I find what third-down plays' realized values as of the next situation would have been if the plays had taken place on fourth down.

Second, the value of going for it depends not only on the team's position on the field, but also on the number of yards it has to go for a first down or touchdown. If there were no need to smooth the estimates, one could use averages to estimate the value of going for it for a specific position and number of yards to go. That is, one could consider all cases where the corresponding circumstance occurred on third down, find what the plays' realized values would have been if they

---

done under the assumption that the differences between the realized and expected values of kicks (the u's in (5)) are uncorrelated with the errors in estimating the V_i's. Although this assumption will not be strictly correct, it is almost certainly an excellent approximation.

had been fourth-down plays, and average the values. In fact, however, there are over a thousand different cases in the sample. Smoothing the estimates is therefore essential.

To smooth the estimates, I focus on the difference between the values of going for it and of turning the ball over on the spot rather than estimating the value of going for it directly. In general, this difference depends on three factors. The first is the difference between the values of having a first down on the spot and of the other team having a first down there. Since the V's are essentially symmetric around the 50-yard line, this factor is essentially independent of the team's position on the field. The second factor is the probability that the team succeeds when it goes for it. As long as the team is not close to its opponent's goal line, there is no reason for this probability to vary greatly with the team's position. The third (and least important) factor is the additional benefit from the yards the team typically gains when it goes for it. Again, as long as the team is not close to the opponent's goal line, there is no reason for the average number of yards it gains to vary substantially with its position. And because the V's are close to linear, the benefit from gaining a given number of yards varies little with position. Thus the third factor is also likely to be almost independent of the team's position over most of the field.

Close to the opponent's goal line, however, the team has less room to work with, and so its chances of success and average number of yards gained are likely to be lower. On the other hand, because the value of a touchdown is much larger than the value of a first down on the 1, the additional benefit from gaining yards may be higher. Thus near the goal line, we cannot be confident that the difference between the values of going for it and of turning the ball over does not vary substantially with the team's position.

Estimating the value of going for it as a fairly general smooth function of the team's position and number of yards to go yields results consistent with this discussion.[11] The resulting estimates of the difference between the values of going for it and of turning the ball over show no

---

[11] Specifically, I divide the field at the 9's, 17's, 33's, and 50. Within each section, I let the value of going for it be quadratic in the team's position and number of yards to go, as in (7) below. Where two sections meet, I constrain both the level of the value of going for it and its derivative with respect to the team's position to be equal for all values of the number of yards to go.

large or systematic variation with the team's position until it is close to the opponent's goal line. For example, the estimate of this difference with 5 yards to go falls from 2.00 at the team's 15 to 1.86 at the 26, rises to 2.25 at the opponent's 43, then falls to 1.97 at the 15. These variations are small, imprecisely estimated, and non-monotonic in a way that does not seem plausible. Thus the most likely possibility is that they largely reflect sampling variation. Starting around the opponent's 15, however, the estimated difference falls more rapidly, reaching a low of 1.10 at the 8.[12]

Because of the large number of parameters that must be estimated with the general approach, the values of going for it in each circumstance are estimated quite imprecisely. Estimating the difference between the values of going for it and turning the ball over provides a convenient way of imposing plausible restrictions. I therefore focus on this difference.

Specifically, let G_{iy} denote the value of going for it on yard line i with y yards to go. The difference between the values of going for it and of turning the ball over on the spot is $G_{iy} - (-V_{i'})$, or $G_{iy} + V_{i'}$, where i' denotes the yard line "opposite" yard line i. From the team's goal line to the opponent's 17, I assume that this difference is independent of i and quadratic in y:

$$G_{iy} + V_{i'} = a_0 + a_1 y + a_2 y^2. \qquad (6)$$

From the opponent's 17 to their goal line, I let the difference depend quadratically on both i and y:

$$G_{iy} + V_{i'} = b_0 + b_1 y + b_2 i + b_3 y^2 + b_4 yi + b_5 i^2 + b_6 y^2 i + b_7 yi^2 + b_8 y^2 i^2. \qquad (7)$$

At the 17, where the two functions meet, I constrain both their level and their derivative with

[12] The estimates also fall relatively rapidly as the team approaches its own goal line. But since a team cannot have only a small number of yards to go when it is close to its goal line, the value of going for it close to the team's goal line is not relevant to the analysis of teams' decisions. In the estimation below, I therefore do not allow for the possibility that the difference changes as the team gets close to its goal line.

respect to i to be equal for all y. This creates 6 restrictions.

The data consist of all third-down plays in the first quarter; there are 4733 observations.[13] Figure 4 summarizes the results. The solid line shows the estimates of $G_{iy} + V_{i'}$ as a function of y for a generic position on the field not inside the opponent's 17, and the dashed line shows the estimates at the opponent's 5. Outside the opponent's 17, the estimate of $G_{iy} + V_{i'}$ for a team facing fourth and 1 is 2.64. On third-and-1 plays from the goal line to the opponent's 17, teams are successful 64 percent of the time, and they gain an average of 3.8 yards; this corresponds to an expected value of 2.66 points.[14] Thus the estimate of 2.64 is reasonable. The estimated difference falls roughly linearly with the number of yards to go. It is 2.05 with 5 yards to go (equivalent to a 45 percent chance of success and an average gain of 6.3 yards), 1.49 with 10 yards to go (a 30 percent chance of success and an average gain of 6.6 yards), and 1.08 with 15 yards to go (an 18 percent chance of success and an average gain of 7.7 yards). These estimates are similar to what one would obtain simply by looking at the average results of the corresponding types of plays.

At the opponent's 5, the estimate of $G_{iy} + V_{i'}$ with 1 yard to go is slightly higher than the estimate elsewhere on the field; it is 2.94, which is equivalent to a 38 percent chance of a first down with an average gain of 2 yards plus a 25 percent chance of a touchdown. The estimate falls more rapidly with the number of yards to go than elsewhere on the field, however. With 5 yards to go, it is 1.42 (equivalent to a 26 percent chance of a touchdown). The estimate for 5 yards to go is quite similar to what one would obtain by looking at averages; the estimate for 1 yard to go is somewhat higher, however.[15]

---

[13] Again, there are a few minor issues with the data. First, to parallel the analysis of kicking, plays that are blown dead before the snap where it would not have been possible to determine if the kicking team had been sent in are excluded. Second, if a penalty causes one third-down play to immediately follow another, both are included. And third, to prevent outliers that are not relevant to decisions about going for it from affecting the results, plays where the team had more than 20 yards to go are excluded.

[14] This translation of average outcomes into point values, and the analogous ones in the rest of the paragraph, are done for a team at midfield. Since the V's are not exactly symmetric around the 50 or exactly linear, choosing a different position would change the calculations slightly.

[15] There are relatively few short yardage third-down and fourth-down plays near the opponent's

The dotted lines show the two-standard-error bands. For the range where $G_{iy} + V_{i'}$ is constrained to be independent of i, the standard errors are small; for 15 or fewer yards to go, they are less than 0.1. Inside the 17, where fewer observations are being used, they are larger, but still typically less than 0.2.

Recommended Choices. Figure 5 combines the analyses of kicking and going for it by showing the number of yards to go where the average payoffs to kicking and going for it are equal as a function of the team's position. On the team's own half of the field, going for it is better on average as long as there are less than about 4 yards to go. After midfield, the gain from kicking falls, and so the critical value rises. It is 6.5 yards at the opponent's 45 and peaks at 9.8 on the opponent's 33. As the team gets into field-goal range, the critical value falls rapidly; its lowest point is 4.0 yards on the 21. Thereafter, the value of kicking changes little while the value of going for it rises. As a result, the critical value rises again. The analysis implies that once a team reaches its opponent's 5, it is always better off on average going for it. The two dotted lines in the figure show the two-standard-error bands for the critical values.[16] The critical values are estimated fairly precisely.

Although these findings contradict the conventional wisdom, they are quite intuitive. As described in the introduction, one case where one can see the intuition clearly is fourth and goal on the 2. The expected payoffs in terms of immediate points to the two choices are very similar, but trying for a touchdown on average leaves the other team in considerably worse field position. Thus trying for a touchdown is better on average. Another case where one can see the intuition fairly easily is fourth and 3 or 4 on the fifty. If the team goes for a first down, it has about a fifty-fifty chance of success; thus both the team and its opponent have about a 50 percent chance of a first and 10. But the team will gain an average of about 6 yards on the fourth-down play; thus it is

goal line where the team can get a first down before the goal line. In addition, the exact estimates of the value of going for it for very short yardages do not affect the conclusions about when teams are on average better off going for it.

[16] For example, the lower dotted line shows the point where the difference between the estimated values of going for it and kicking is twice its standard error.

on average better off than its opponent if it goes for it. If the team punts, its opponent will on average end up with a first and 10 around its 14. Both standard views about football and the analysis in Section II suggest that the team and its opponent are about equally well off in this situation. Thus, the team is on average better off than its opponent if it goes for a first down, but not if it punts. Going for the first down is therefore preferable on average.

The very high critical values in the dead zone also have an intuitive explanation. The chances of success if the team goes for it decline only moderately as the number of yards to go increases. For example, away from the opponent's end zone, teams obtain a first down or touchdown on third down 64 percent of the time when they have 1 yard to go, 44 percent of the time when they have 5 yards to go, and 34 percent of the time when they have 10 yards to go. As a result, the value of going for it falls only moderately as the number of yards to go rises. Thus the large decrease in the gain from kicking in the dead zone causes a large increase in the critical value.

Actual Choices. Teams' actual choices are dramatically more conservative than the choices recommended by the dynamic-programming analysis. On the 1604 fourth downs in the sample where the analysis implies that teams are on average better off kicking, they went for it only 9 times. But on the 1068 fourth downs where the analysis implies that teams are on average better off going for it, they kicked 959 times.[17]

The dashed line in Figure 5 summarizes teams' actual choices. It shows, for each point on the field, the largest number of yards to go with the property that when teams have that many or fewer yards to go, they go for it at least as often as they kick. Over most of the field, even with 1 yard to go teams usually kick. Teams are slightly more aggressive in the dead zone, but are still far less aggressive than the dynamic-programming analysis suggests. On the line summarizing teams' actual choices, the null hypothesis that the average values of kicking and going for it are

[17] These figures exclude the 28 cases where we cannot observe the team's intent because of a penalty before the snap.

equal is typically rejected with a t-statistic between 3 and 7.[18]

IV. COMPLICATIONS

This section discusses six considerations that have been omitted from the basic analysis. The first two concern the estimation of the V's, the next three concern the choice between kicking and going for it given the V's, and the final one concerns both issues.

Rational Risk Aversion. I have assumed that teams are risk-neutral concerning points scored. But since teams' goal is to maximize the probability of winning, this is not correct: to some extent, what is important is scoring some points, not scoring a large number. The analysis may therefore overstate the value of a touchdown relative to a field goal, and thus overstate the benefits of going for it on fourth down.

Three considerations suggest that this effect is not important. First, it is largely irrelevant to decisions in the middle of the field. Near midfield, a team should maximize the probability that it

[18] Carter and Machol (1978) and Carroll, Palmer, and Thorn (1998, Ch. 10) also examine fourth-down decisions. Carter and Machol only consider decisions inside the opponent's 35-yard line. They use estimates from their earlier work (described in n. 6 above) to assign values to different situations. To estimate the payoff from going for it, they pool third-down and fourth-down plays. They assume that all successful plays produce exactly the yards needed for a first down and that all unsuccessful plays produce no yards, and that the probability of success does not depend on the team's position on the field. They then compare the estimated payoffs to going for it with the payoffs to field-goal attempts and punts. They conclude that teams should be considerably more aggressive than they are.

Carroll, Palmer, and Thorn consider decisions over the entire field. They do not spell out their method for estimating the values of different situations (though it appears to be related to Carter and Machol's), and it yields implausible results. Similarly to Carter and Machol, they pool third-down and fourth-down plays, assume that successful plays produce one more yard than needed for a first down and that unsuccessful plays yield no gain, and assume that the chances of success do not vary with field position. They again conclude that teams should be considerably more aggressive. Their specific findings about when going for it is preferable on average are quite different from mine, however.

Finally, neither Carter and Machol nor Carroll, Palmer, and Thorn investigate the statistical significance of their results.

is the first to get close to the opponent's goal line, since that is necessary for either a field goal or a touchdown. But teams are conservative over the entire field.

Second, as I show in Section V, teams are conservative even in situations where the behavior that would maximize the chances of winning would be risk-loving over points scored.

Third, it is possible to obtain some direct evidence about the impact of points on the probability of winning. Because teams adjust their play late in the game according to the score, one cannot just look at the distribution of actual winning margins. For example, a team that is trailing by six points late in the game will try for a touchdown, a team that is trailing by two will try for a field goal, and a team that is leading by one will try to run out the clock. As a result, looking at the fraction of games that are decided by one point will lead to an overestimate of the value of an additional point early in the game.

Instead, I try to approximate what the distribution of winning margins would be in the absence of late-game adjustments, and use this to estimate the value of a field goal or touchdown early in the game. I begin by dividing the games into deciles according to the point spread. I then find the score for the favorite and the underdog at the end of the first half; the idea here is that these scores are relatively unaffected by adjustments in response to the score. I then construct synthetic final scores by combining the first-half scores of each pair of games within a decile. This yields a total of 74(73)/2 (or 73(72)/2) synthetic games for each decile, for a total of 26,718 observations. I use the results to estimate the impact of an additional field goal or touchdown in the first quarter. For example, the estimated effect of a field goal on the probability of winning is the sum of the probability that a team would trail by one or two points at the end of the game plus half the probability that the score would be tied or the team would trail by three.

This exercise suggests that 7 points are in fact slightly <u>more</u> than seven-thirds as valuable as 3. 3 additional points are estimated to raise the probability of winning by 6.8 percentage points; 7 additional points are estimated to raise the probability of winning by 16.2 percentage points, or 2.40 times as much. The source of this result is that the distribution of synthetic margins is considerably higher at 4 and 7 points than at 1 or 2 points. To put it differently, to some extent what is important about a touchdown is not that its usual value is 7 points, but that its usual value is between two and three times the value of a field goal.

Selection Bias. Teams are not assigned to situations randomly. For example, good teams are more likely to have first downs near their opponents' goal line, and poor teams are more likely to have first downs near their own goal line.

It is not clear how this might bias the analysis. It appears likely to lead to some overestimate of the value of a first down near the opponent's goal line, and thus of the incentive to be aggressive in the opponent's territory. But it also appears likely to lead to some overestimate of the value of field position, and thus of the value of punting.

To examine the importance of this potential difficulty, I reestimate the V's excluding all observations from the third of games where the point spread is greater than or equal to seven. This change reduces the heterogeneity in the sample considerably, and thus is likely to reduce any effects of selection bias.

The change in the sample has only a small impact on the estimates of the V's. Except on a team's 1 and 2 yard lines, the estimates never change by more than 0.08. Equally important, the changes are not in the direction one would expect if selection bias were important: the estimated V's rise rather than fall near the opponent's goal line, and they do not change in any consistent way near the team's own goal line. Thus the results provide no evidence that selection bias is important.

Third Down versus Fourth Down. The analysis uses the outcomes of third-down plays to gauge what would happen if teams went for it on fourth down. There are two ways to investigate whether this is likely to introduce important bias. The first is to consider how teams' incentives are likely to affect outcomes on fourth downs relative to third downs. The second is to directly compare outcomes on the two downs.

To consider incentives, note that the relative payoffs to different outcomes are different on fourth down than on third down. In particular, the benefit from a long gain relative to just making a first down is smaller on fourth down. As a result, both the offense and 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 just making a first down, while the defense will be willing to do the reverse.

It follows that one would expect the direction of the bias from using third-down plays to depend on which team has more influence on the distribution of outcomes. For example, consider the extreme case where the offense always calls the same play but the defense can adjust its behavior. In this case, the defense can do at least as well as it would if the distribution of outcomes on fourth downs were the same as on third downs, and it may be able to do better. Thus using third downs to gauge what would happen on fourth downs would lead to overestimates of the value of going for it. Similarly, if only the offense can influence the distribution of outcomes, using third downs would lead to underestimates of the value of going for it. Since it seems unlikely that the defense has substantially more scope than the offense to affect the distribution of outcomes, this suggests that the use of third downs is unlikely to lead to substantial overestimates of the value of going for it.

More importantly, the relative payoffs to different outcomes do not differ greatly between third and fourth downs. This is clearest for cases where the team has third and goal or fourth and goal. To a first approximation, on either down the offense will try to maximize its chances of a touchdown and the defense will try to minimize it. Thus both sides' behavior on fourth down should be essentially the same as on third down. But even away from the goal line, the relative payoffs to different outcomes are similar on third and fourth down. For example, consider a team that is on its 30 and needs 2 yards for a first down. On third down (under the realistic assumption that the team will punt if it fails to make a first down), the benefit of gaining 15 yards rather than none is 1.4 times as large as the benefit of gaining 2 yards rather than none. On fourth down, the benefit of gaining 15 yards rather than none is 1.2 times as large as the benefit of gaining 2 yards rather than none. Thus, one would not expect teams to behave very differently on the two downs. As a result, any bias from the use of third-down plays is likely to be small.

The second approach to investigating the appropriateness of using the outcomes of third-down plays is to examine how closely outcomes on fourth downs resemble those on third downs.

In particular, one can compare the realized values of plays where teams went for it on fourth downs (that is, the immediate points scored plus the value of the resulting field position) with what one would expect based on the analysis of third downs.

This comparison is potentially problematic, however, for two reasons. First, the sample size is small. As described above, teams went for it only 118 times in the sample, whereas there are 4733 third-down plays. Second, times when teams choose to go for it are likely to be unusual: the teams may know that they are particularly likely to succeed, or they may be desperate.

To increase the sample without bringing in fourth-down attempts that are likely to be especially unusual, I expand the sample to include the entire game except for the last two minutes of each half (and overtimes). This increases the sample to 1338 plays. And as a partial remedy for the second problem, I experiment with controlling for the amount the team with the ball is trailing by and the amount it is favored by.

The results suggest that fourth downs are virtually indistinguishable from third downs. The mean of the difference between the realized value of the fourth-down attempts and what is predicted by the analysis of third downs is 0.006 (with a standard error of 0.7), which is essentially zero. When controls for the prior point spread and the current point differential are included, the coefficient falls to -0.042 and remains highly insignificant. The point estimate corresponds to the probability of success being one percentage point lower on fourth downs than on third downs, which would have almost no impact on the analysis.

Additional Information. In choosing between kicking and going for it, a team can use more information than the averages employed in the dynamic-programming analysis. It has information about the quality of its offense and the opponent's defense, the quality of its punter and placekicker and the opponent's punt returner, the weather, and so on. Thus it would not be optimal for it to follow the recommendations of the dynamic-programming analysis mechanically.

Additional information cannot, however, account for the large systematic departures from the recommendations of the dynamic-programming analysis. Over wide ranges, teams almost always kick in circumstances where the analysis implies that they would be better off on average going for it. For example, on the 512 fourth downs in the sample in the offense's half of the field where the dynamic-programming analysis suggests going for it, teams went for it only 7 times. Similarly, on the 175 fourth downs with 5 or more yards to go where the analysis suggests going for it, teams went for it only 13 times.

Additional information can account for this behavior only if the information takes the form of teams knowing on a large majority of fourth downs that the expected payoff to going for it relative to kicking is considerably less than average, and knowing on the remainder that the expected payoff is dramatically larger than average. This possibility is not at all plausible. Further, it predicts that when teams choose to go for it, the results will be far better than one would expect based on averages. As described above, this prediction is contradicted by the data.

Momentum. Failing on fourth down could be costly to a team's chances of winning not just through its effect on possession and field position, but also through its effect on energy and emotions. As a result, it might be more costly for the other team to have the ball as a result of stopping a fourth-down attempt than for it to have the ball at the same place on the field in the course of a normal drive or as the result of a punt. In this case, the analysis would understate the cost of a failed fourth-down attempt.

There are two reasons to be skeptical of this possibility. First, the same reasoning suggests that there could be a motivational benefit to succeeding on fourth down, and thus that the analysis could understate the gain from a successful fourth-down attempt. Second, studies of momentum in other sports have found at most small momentum effects (see, for example, Gilovich, Vallone, and Tversky, 1985; Albright, 1993; and Klaassen and Magnus, 2001).

More importantly, it is possible to obtain direct evidence about whether outcomes differ systematically from normal after plays whose outcomes are either very bad or very good. To obtain a reasonable sample size, I do not look only at fourth-down attempts. Instead, for very bad plays I consider all cases where from one situation to the next, possession changed and the ball advanced less than ten yards. This captures not only failures on fourth downs, but also many turnovers, failed field goals, blocked punts, and long punt returns. For very good plays, I simply consider all cases where the offense scored a touchdown. These criteria yield 636 very bad plays and 628 very good plays. I then examine what happens from the situation immediately following the extreme play to the next situation, from that situation to the next, and from that situation to the subsequent one. In each case, I ask whether the realized values of these situations one situation later differ systematically from the V's for those situations. That is, I look at the means of the relevant e_{gt}'s (always computed from the perspective of the team that had the ball before the very bad or very good play).

The results provide no evidence of momentum effects. All the point estimates are small and highly insignificant; the largest t-statistic (in absolute value) is less than 1.3. Moreover, the largest point estimate (again in absolute value) goes the wrong direction from the point of view of the momentum hypothesis: from the situation immediately following a very bad play to the next, the team that lost possession does somewhat better than average.

Partial Equilibrium versus General Equilibrium. The analysis looks at decisions on individual fourth downs, taking all other decisions as given. But these decisions could affect other choices.

If both teams follow the recommendations of the dynamic-programming analysis, their offenses will do better on average. This suggests that the value of having the ball anywhere on the field will be greater than the partial-equilibrium estimates imply. In this case, the benefits of going for it relative to kicking will be even larger than the preceding analysis suggests, particularly when teams are far from their opponents' goal line.

If only one team is more aggressive on fourth downs, on the other hand, it will on average score more points; this implies that any situation (including ones where its opponent has the ball) will be more valuable to it. There is no evident reason for this to have an important effect on optimal choices. It appears likely, however, that the increase in value will be larger in situations where the team is more likely to face a fourth down soon, such as when it has the ball in near its own goal line. If this is correct, it implies that the analysis tends to understate the value of going for it near one's goal line.

We will see in Section V, however, that there are relatively few plays where the analysis suggests that teams should behave differently than they do. Thus, these effects are probably small. Moreover, for the most part this discussion suggests that the analysis understates the gap between actual and optimal strategies.

More aggressive choices on fourth downs could change not only the V's, but also behavior on other downs. Most importantly, greater aggressiveness by a team on fourth downs (regardless of whether its opponent is also more aggressive on fourth downs) raises the relative payoff to being stopped just short of a first down or touchdown on third down. Both the offense and defense will adjust their behavior in response to this change in relative payoffs. As with the use of third-down plays to estimate the likely outcomes of fourth-down plays, if the offense has more influence than the defense over the distribution of outcomes, this will on net benefit the offense. In this case, the estimates understate the benefits of improved decisions on fourth downs. If the defense has more influence on the distribution of outcomes, the opposite is true.

Again, however, the fact that there are not many plays where the analysis suggests that teams should act differently makes it unlikely that the spillover effects to third down are large. In addition, calculations like those for third down versus fourth down suggest that the impact on relative payoffs is not large, and thus again that the effects are likely to be small. Finally, since it seems unlikely that the defense has substantially more influence on the distribution of outcomes than the offense, it is unlikely that omitting general-equilibrium effects results in substantial overestimates of the value of improved fourth-down decisions.

V. QUANTITATIVE IMPLICATIONS

An obvious question is whether the potential gains from different choices are important. There are in fact two distinct questions. The first is whether there are cases of clear-cut departures from win-maximization. If there were not, then small changes in the analysis might reverse the conclusions, which would suggest that the evidence that teams are not maximizing their chances of winning is not strong.

The answer to this question is that there are clear-cut departures. One example is the case of fourth and goal on the 2 discussed above. The estimates imply that trying for a touchdown and failing is only slightly worse than kicking a field goal. As a result, they imply that going for a touchdown is preferable on average as long as the probability of success is at least 18 percent, while the actual probability of success is about 45 percent. Thus there are no plausible changes in the analysis that could reverse the conclusion that trying for a touchdown is preferable on average. Moreover, the average benefit of trying for a touchdown in this case is substantial. The estimated value of going for it (the relevant G_{iy}) is roughly 3.7 points, while the estimated value of kicking (K_i) is about 2.4 points. Since each additional point raises the probability of winning by about 2.3 percentage points, trying for a touchdown on average increases the team's chances of winning by about 3 percentage points. Yet teams attempted a field goal every time in the sample that they were in this position.

Two other examples are fourth and goal on the 1 and fourth and 1 between the opponent's 35 and 40. For the first, the estimates imply that the critical and actual probabilities of success are 16 percent and 62 percent, and that trying for a touchdown on average increases the chances of winning by about 5 percentage points. For the second, the critical and actual probabilities are 39 percent and 64 percent, and going for a first down raises the probability of winning by about 2.5 percentage points. In these cases, teams do not always kick, but they do about half the time. These decisions are consistent with win-maximization only if teams have substantial additional information that allows them to identify times when their fourth-down attempts are especially likely to succeed. As described in the previous section, there is no evidence of such large additional information. Moreover, if teams did have such information, it would mean that some of their decisions in the cases where they virtually always kick represent even larger departures from win-maximizing choices.

The second question is whether the analysis implies that teams could increase their overall chances of winning substantially. Since the analysis considers only a small fraction of plays and only a single decision on those plays, one would not expect it to show large potential increases in the chances of winning. And indeed, the overall potential gains are small. In the 732 first quarters in the sample, there are 959 cases where a team kicked when the difference between the estimated values of going for it and kicking was positive, or an average of 0.66 cases per team-quarter. The average estimated value of the expected gain from going for it in these cases is 0.35 points. Thus the expected payoff to a typical team of being more aggressive on fourth downs in the first quarter is approximately 0.23 points per game, which corresponds to an increase in the probability of winning of about 0.5 percentage points.

Teams could also benefit by being more aggressive on fourth downs in the remaining quarters. A full-fledged analysis of fourth-down decisions over the entire game would require accounting for the score and the time remaining, which would complicate the analysis enormously. Nonetheless, the evidence is clear. Consider first all fourth-down plays in the second, third, and fourth quarters. In the 3555 cases where the analysis of decisions in the first quarter suggests that going for it is preferable on average, teams went for it only 966 times. And in the 5678 cases where that analysis suggests that kicking is preferable, they went for it only 460 times. That is, teams are almost as conservative over the last three quarters as they are in the first.[19] But there is no reason to think the average benefits to various outcomes are much different in the later quarters.

Stronger evidence comes from restricting the sample to cases where win-maximization implies that teams should be risk-loving over points scored (and where they are not so far behind that they might reasonably view the game as unwinnable). Specifically, I consider fourth downs in the second quarter where the team with the ball is trailing by at least 4 points; fourth downs in the third quarter where the team is trailing by between 4 and 28 points; and fourth downs in the fourth quarter where the team is trailing by between 4 and 16 points. In the 3065 such cases in the sample, the first-quarter analysis suggests going for it 1147 times, yet teams went for it only 596

[19] As before, these figures (and those in the next paragraph) exclude cases where it is not possible to determine the team's intent because of a penalty before the snap.

times. That is, in cases where win-maximizing behavior is risk-loving over points scored, teams are considerably more conservative than they would be if they acted in a way that was risk-neutral over points.

This evidence suggests that a rough estimate of the potential gains from going for it more often on fourth downs over the whole game is four times the gains from the first quarter, or an increase of about 2.1 percentage points in the probability of winning. Since an NFL season is 16 games long, this corresponds to slightly more than one additional win every three seasons. This is a modest (though not trivial) effect. Thus, one cannot rule out the possibility that I have merely identified a clear-cut but modest and isolated departure from maximization. Because I have examined one particular type of decision in detail, there is simply no evidence either for or against this hypothesis.

VI. CONCLUSION

Much economic analysis is built on the idea that the assumption that agents maximize simple objective functions leads to reasonably accurate descriptions of their behavior. This paper demonstrates that in a case where this hypothesis can be tested directly, it fails. Specifically, the paper shows that the behavior of National Football League teams on fourth downs departs systematically from the behavior that would maximize their chances of winning. The departure is large and overwhelmingly statistically significant, and it cannot be explained by rational risk aversion, information known to teams that is omitted from the analysis, or other complications. This is true even though the decisions are comparatively simple, the possibilities for learning and imitation are unusually large, the compensation for the coaches who make the decisions is extremely high, and the market for their services is intensively competitive. Despite these forces, the coaches who fail to make maximizing choices are not fired and replaced by ones who do. How can this be?

The departures from maximizing behavior are toward "conservative" behavior: the immediate payoff to a punt or field-goal attempt has a lower variance than the immediate payoff to going for it. Nonetheless, conventional risk aversion cannot explain the results. At the end of the game, one team will have won and the other will have lost. Thus even a decision-maker who faces a large cost of losing and little benefit of winning should maximize the probability of winning.

At a broad level, two forces could lead to systematic departures from win-maximizing behavior toward conservative choices.[20] First, the relevant actors could have preferences not just over whether their team wins or loses, but over the probability of winning during the game, and they could be risk-averse over this probability. That is, individuals may value decreases in the probability of winning as the result of failed gambles and increases as the result of successful gambles asymmetrically. Such risk aversion over probabilities could come from fans, owners, or coaches and players. If it comes from fans (and if their risk aversion affects their demand), teams' choices would be departures from win-maximization but not from profit-maximization. If it comes from owners, then they would be forgoing some profits to obtain something else they value. And if it comes from coaches and players, the departures from win-maximization could again be profit-maximizing (if coaches and players are willing to accept lower compensation to follow more

---

[20] The departures from win-maximizing choices need not stem entirely from forces promoting conservatism. At least two other factors could play some role. First, the departures could be partly the result of history: optimal choices may have been more conservative in the past, when offenses were less potent, and decision-makers may not have adapted fully to the changes in the game. But Carter and Machol (1978) present evidence that the tendency toward conservative decisions on fourth downs was present 30 years ago; thus only extremely slow adjustment could explain the results in this way.

Second, decision-makers may tend to overweight points scored quickly; this could account for some of teams' tendency to attempt field goals when going for it is preferable and some of their extreme reluctance to go for it deep in their own territories. But this cannot explain decisions to punt near mid-field and slightly into the opponent's territory; yet the departures from win-maximization appear roughly as large in these cases as elsewhere on the field.

In addition, herding (for example, Scharfstein and Stein, 1990) could magnify departures from win-maximization. That is, if coaches who deviate from standard practice are punished more for failures than they are rewarded for successes, departures from win-maximization will be self-reinforcing. Herding cannot explain why the departures are in one particular direction, however.

Thus, it appears that a full explanation of the departures from win-maximization must assign an important role to forces leading to conservative behavior.

conservative strategies); or they could be the result of agency problems (if coaches and players have some ability to make choices that are not profit-maximizing).

The amount of risk aversion over probabilities needed to account for the results is very large, however. For example, consider a team facing fourth and 1 between its opponent's 35 and 40 early in the game. In this position, teams go for it about half the time. The estimates suggest that if the team punts, its chances of winning the game are about 51 percent. If it goes for a first down, it has about a 64 percent chance of success, in which case its chances of winning are about 57 percent, and a 36 percent chance of failure, in which case its chances of winning are about 47 percent; the implied overall chance of winning is 53.3 percent. Thus, by punting the team is gaining only a small reduction in variance at a substantial cost in terms of the probability of winning. To be indifferent between the two choices, a decision-maker must place roughly three times as much weight on reductions in the probability of winning below what can be attained by the punt as on increases in the probability of winning from that level.

The second potential source of systematically conservative decisions involves a more significant departure from standard models. The individuals involved may want to maximize the chances of winning, but do so imperfectly. Many skills are more important to running a successful football team than a command of mathematical and statistical tools. And it would hardly be obvious to someone without a knowledge of those tools that mathematical and statistical analysis would have any significant value to a football team. Thus the decision-makers may want to maximize their teams' chances of winning, but rely on experience and intuition rather than formal analysis. And because individuals are risk-averse in other contexts, experience and intuition may lead them to behave more conservatively in this context than is appropriate for maximizing their chances of winning. Likewise, their risk aversion in other contexts could cause failed fourth-down attempts to be particularly memorable or salient to them, and so cause them to underestimate the chances of success on fourth down. As with risk aversion over probabilities, such imperfect optimization would have to be substantial to account for the results, however.

The experimental and behavioral literatures have documented many instances of behavior that is systematically more conservative than standard models predict. In the classic "Ellsberg paradox" (Ellsberg, 1961), individuals act as though they are risk-averse over probabilities. More recent studies also consistently find evidence of such behavior in contexts where probabilities are ambiguous (for example, Hogarth and Kunreuther, 1989).

Selten, Sadrieh, and Abbink (1999) present even stronger evidence of departures from risk neutrality over probabilities. They report the results of experiments where subjects face choices among lotteries whose payoffs are themselves lottery tickets. They find that changing the payoffs from money to lottery tickets exacerbates departures from risk neutrality and from the predictions that expected utility theory makes concerning monetary gambles. For cases where subjects have often been found to act risk-loving over monetary payoffs, they also often act risk-loving over probabilities. But for gambles that are most similar to fourth-down decisions (in the sense that there are no low-probability, high-stakes outcomes), subjects exhibit considerable risk aversion over probabilities.[21]

Another strand of work shows that individuals tend to make conservative decisions concerning monetary gambles in ways that cannot be rationalized by any plausible degree of risk aversion (Rabin, 2000). And there is considerable evidence that both individuals and firms tend to view risky decisions in isolation, and to exhibit risk aversion regarding them even when their implications for the risk of the individual's overall wealth or the firm's total profits are minimal (for example, Kahneman and Lovallo, 1993; Read, Loewenstein, and Rabin, 1999).

Much of the existing evidence of systematically conservative behavior involves highly stylized laboratory settings with small stakes and inexperienced decision-makers devoting relatively little effort to their choices. Thus previous work provides little evidence about the strength of the forces pushing decision-makers toward conservatism. The results of this paper suggest that the forces may be shockingly strong. In a high-stakes environment, the most sought-

[21] Thaler (2000) speculates that evidence from a variety of sports contexts would also show decision-makers acting as though they are risk-averse over probabilities. He does not present systematic evidence in support of this claim, however.

after decision-makers in the industry, faced with decisions that arise repeatedly, make choices that differ systematically and in a quantitatively significant way from the predictions of models of full optimization of simple objective functions. Moreover, they do so in cases (such as the examples at midfield and near the opponent's goal line) where the relevant considerations are not overwhelmingly complex, and where the amount they are giving up in terms of the simple objective function appears substantial relative to any possible offsetting benefits.

Unfortunately, little in the experimental and behavioral literatures bears on the question of whether patterns of conservative behavior arise because individuals have non-standard objective functions or because they are imperfect maximizers. For example, individuals may act as if they are risk-averse over probabilities either because they genuinely dislike uncertainty about probabilities, or because they misapply their usual rules of thumb about the importance of avoiding risk to settings where the risk involves probabilities rather than payoffs. Similarly, as Read, Loewenstein, and Rabin observe, individuals may choose to forgo a sequence of gambles that is virtually certain to have a positive total payoff either because the expected utility from the eventual payoff is not enough to compensate them for the disutility they would suffer from the many small setbacks along the way, or because they do not understand how favorable the distribution of final outcomes would be. In other words, previous work has relatively little to say about the critical issue of whether systematic departures from the predictions of simple models of optimization are the result of different objective functions or imperfect optimization. And as described above, the departures from win-maximization in football could also arise from either source.

The hypotheses of non-standard objective functions and imperfect optimization do, however, make different predictions about the future evolution of football strategy. If choices are conservative because of preferences concerning the probability of winning during the game, behavior will not change. If choices are conservative because of imperfect optimization, on the other hand, then trial-and-error, increased availability of data, greater computing power, and the development of formal analyses of strategy will cause behavior to evolve toward victory-maximizing choices. Thus the future evolution of football strategy will provide evidence about the merits of these two competing views of the source of systematic departures from the predictions of models of complete optimization of simple objective functions.

APPENDIX

This appendix describes the main rules of football that are relevant to the paper.

A football field is 100 yards long. Each team defends its own goal line and attempts to move the ball toward its opponent's. The yard lines are numbered starting at each goal line and are referred to according to which team's goal line they are closer to. Thus, for example, the yard line 20 yards from one team's goal line is referred to as that team's 20-yard line.

The game begins with a kickoff: one team puts the ball in play by kicking the ball from its own 30-yard line to the other team. After the kickoff, the team with the ball has four plays, or downs, to move the ball 10 yards. If at any point it gains the 10 yards, it begins a new set of four downs. Plays are referred to by the down, number of yards to go for a first down, and location. For example, suppose the receiving team returns the opening kickoff to its 25-yard line. Then it has first and 10 on its own 25. If it advances the ball 5 yards on the first play, it has second and 5 on its own 30. If it advances 8 yards on the next play (for a total of 13), it now has first and 10 on its own 38. The team with the ball is referred to as the offense, the other team as the defense.

If a team advances the ball across its opponent's goal line, it scores a touchdown. A touchdown gives the team six points and an opportunity to try for an extra point, which almost always produces one point. If a team has a first and 10 within 10 yards of its opponent's goal line, it cannot advance 10 yards without scoring a touchdown. In this case, the team is said to have first and goal rather than first and 10.

On fourth down, the offense has three choices. First, it can attempt a conventional play. If the play fails to produce a first down or touchdown, the defense gets a first down where the play ends. Second, it can kick (or "punt") the ball to the defense; this usually gives the defense a first down, but at a less advantageous point on the field. Third, it can attempt to kick the ball through the uprights located 10 yards behind the opponent's goal line (a "field goal"). If it succeeds, it scores 3 points. If it fails, the defense gets a first down at the point where the kick was made, which is normally 8 yards farther from its goal line than the play started. (If the field goal was attempted from less than 20 yards from the goal line, however, the defense gets a first down on its 20-yard line rather than at the point of the attempt.) After either a touchdown or field goal, the scoring team kicks off from its 30-yard line, as at the beginning of the game.

The final (and by far the least common) way to score is a safety: if the offense is pushed back across its own goal line, the defense scores 2 points, and the offense puts the ball in play by kicking to the other team from its 20-yard line (a "free kick").

The game is divided into four 15-minute periods. At the beginnings of the second and fourth quarters, play continues where it left off. At the beginning of the third quarter, however, play begins afresh with a kickoff by the team that did not kick off at the beginning of the game.

REFERENCES

Albright, S. Christian. 1993. "A Statistical Analysis of Hitting Streaks in Baseball." Journal of the American Statistical Association 88 (December): 1175-1183.

Alchian, Armen A. 1950. "Uncertainty, Evolution, and Economic Theory." Journal of Political Economy 58 (June): 211-221.

Becker, Gary S. 1957. The Economics of Discrimination. Chicago: University of Chicago Press.

Carroll, Bob, Pete Palmer, and John Thorn, with David Pietrusza. 1998. The Hidden Game of Football: The Next Edition. New York: Total Sports.

Carter, Virgil, and Robert E. Machol. 1971. "Operations Research on Football." Operations Research 19 (March-April): 541-544.

Carter, Virgil, and Robert E. Machol. 1978. "Optimal Strategies on Fourth Down." Management Science 24 (December): 1758-1762.

Chiappori, P.-A., S. Levitt, and T. Groseclose. 2002. "Testing Mixed-Strategy Equilibria When Players Are Heterogeneous: The Case of Penalty Kicks in Soccer." American Economic Review 92 (September): 1138-1151.

Ellsberg, Daniel. 1961. "Risk, Ambiguity, and the Savage Axioms." Quarterly Journal of Economics 75 (November): 643-669.

Fama, Eugene F. 1980. "Agency Problems and the Theory of the Firm." Journal of Political Economy 88 (April): 288-307.

Friedman, Milton. 1953. "The Methodology of Positive Economics." In Essays in Positive Economics, 3-43. Chicago: University of Chicago Press.

Gilovich, Thomas, Robert Vallone, and Amos Tversky. 1985. "The Hot Hand in Basketball: On the Misperception of Random Sequences." Cognitive Science 17 (July): 295-314.

Hogarth, Robin M., and Howard Kunreuther. 1989. "Risk, Ambiguity, and Insurance." Journal of Risk and Uncertainty 2 (April): 5-35.

Kahneman, Daniel, and Dan Lovallo. 1993. "Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking." Management Science 39 (January): 17-31.

Klaassen, Franc J. G. M., and Jan R. Magnus. 2001. "Are Points in Tennis Independent and Identically Distributed? Evidence from a Dynamic Binary Panel Data Model." Journal of the American Statistical Association 96 (June): 500-509.

Rabin, Matthew. 2000. "Risk Aversion and Expected-Utility Theory: A Calibration Theorem." Econometrica 68 (September): 1281-1292.

Read, Daniel, George Loewenstein, and Matthew Rabin. 1999. "Choice Bracketing." Journal of Risk and Uncertainty 19 (December): 171-197.

Scharfstein, David S., and Jeremy C. Stein. 1990. "Herd Behavior and Investment." American Economic Review 80 (June): 465-479.

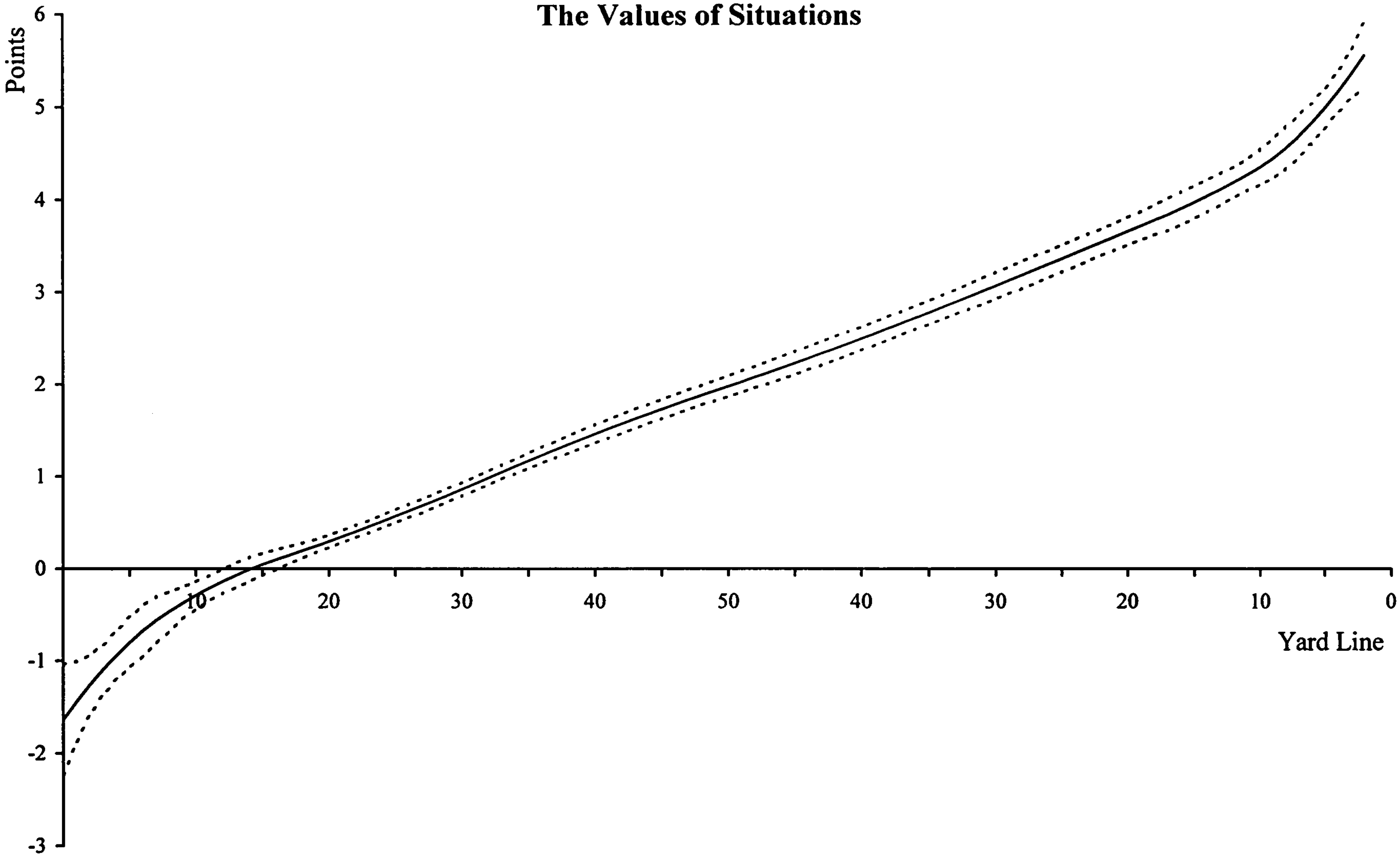
Selten, Reinhard, Abdolkarim Sadrieh, and Klaus Abbink. 1999. "Money Does Not Induce Risk Neutral Behavior, But Binary Lotteries Do Even Worse." Theory and Decision 46 (June): 211-249.

Thaler, Richard H. 2000. "Sudden Death Aversion." Unpublished paper, Graduate School of Business, University of Chicago (November).

Walker, Mark, and John Wooders. 2001. "Minimax Play at Wimbledon." American Economic Review 91 (December): 1521-1538.

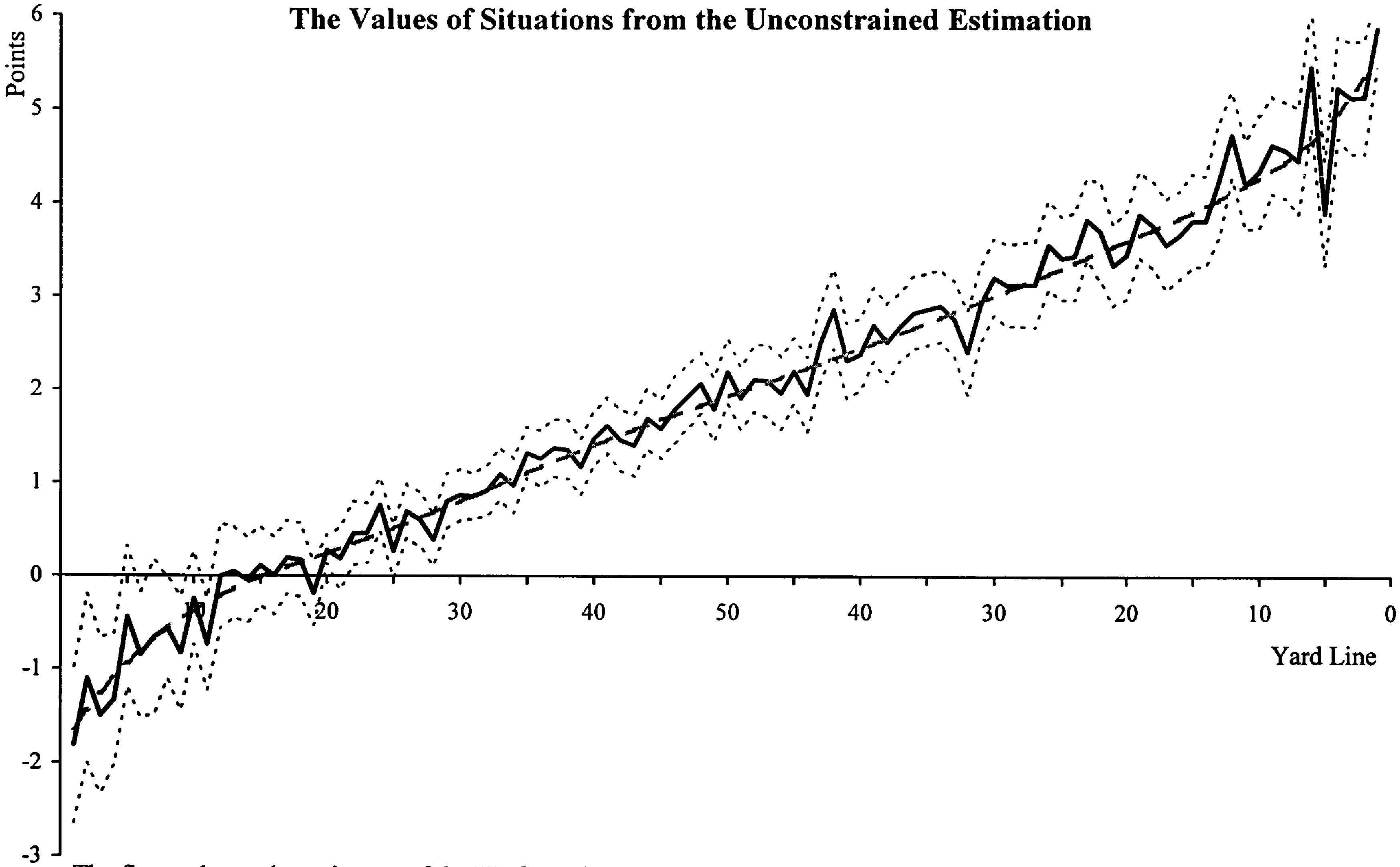
Figure 1

The Values of Situations



The solid shows the estimates of the V's from the splined estimation. The dotted lines show the two-standard-error bands. The estimated value of a kickoff is -0.62 (with a standard error of 0.04); the estimated value of a free kick is -1.21 (0.51).

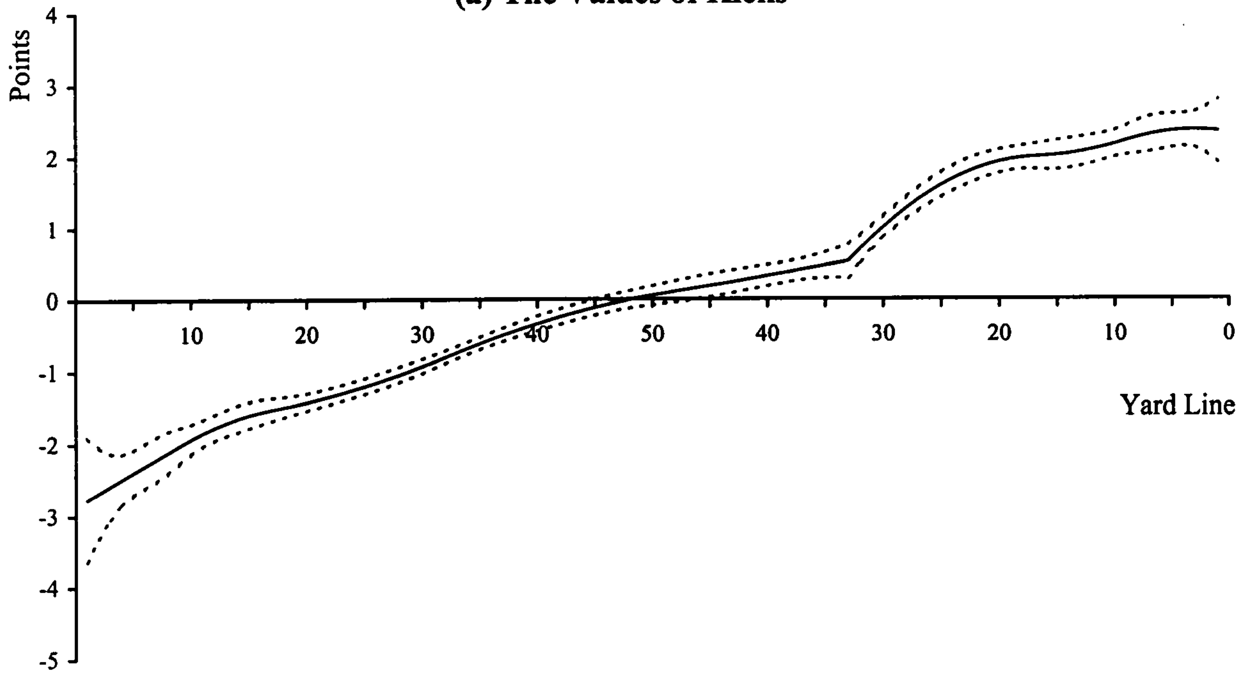
Figure 2



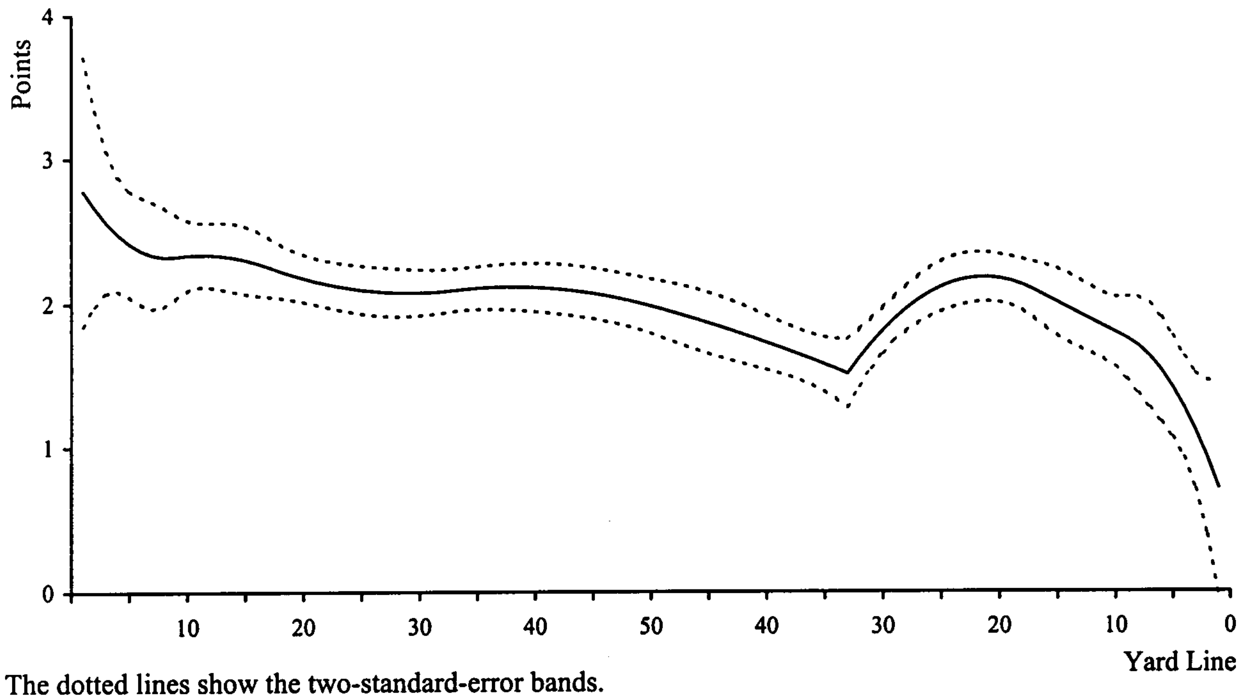
The figure shows the estimates of the V's from the unconstrained estimation. The dotted lines show the two-standard-error bands. The dashed line shows the estimates of the V's from the splined estimation.

Figure 3

(a) The Values of Kicks



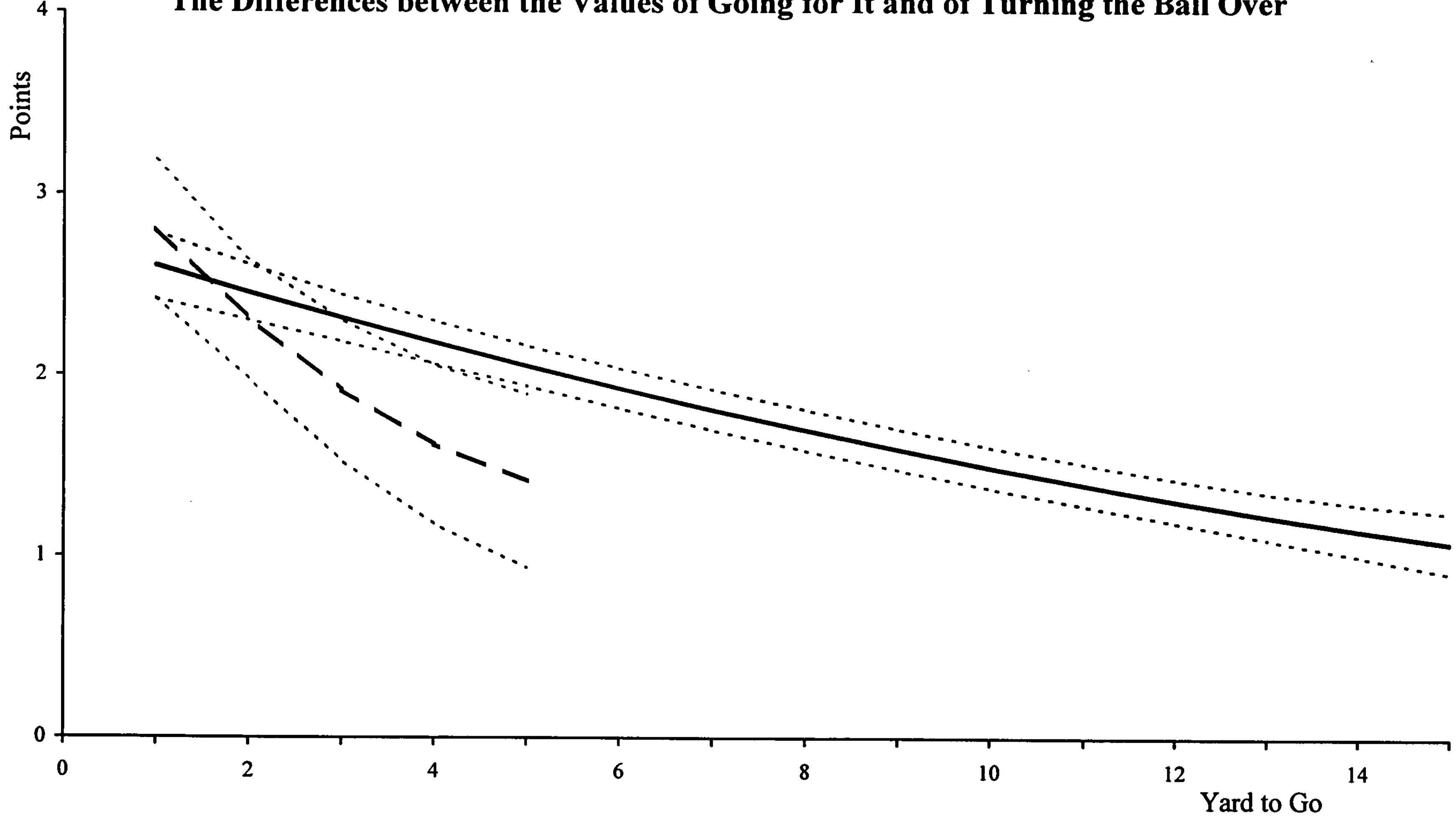
(b) The Differences between the Values of Kicks and of Turning the Ball Over



The dotted lines show the two-standard-error bands.

Figure 4

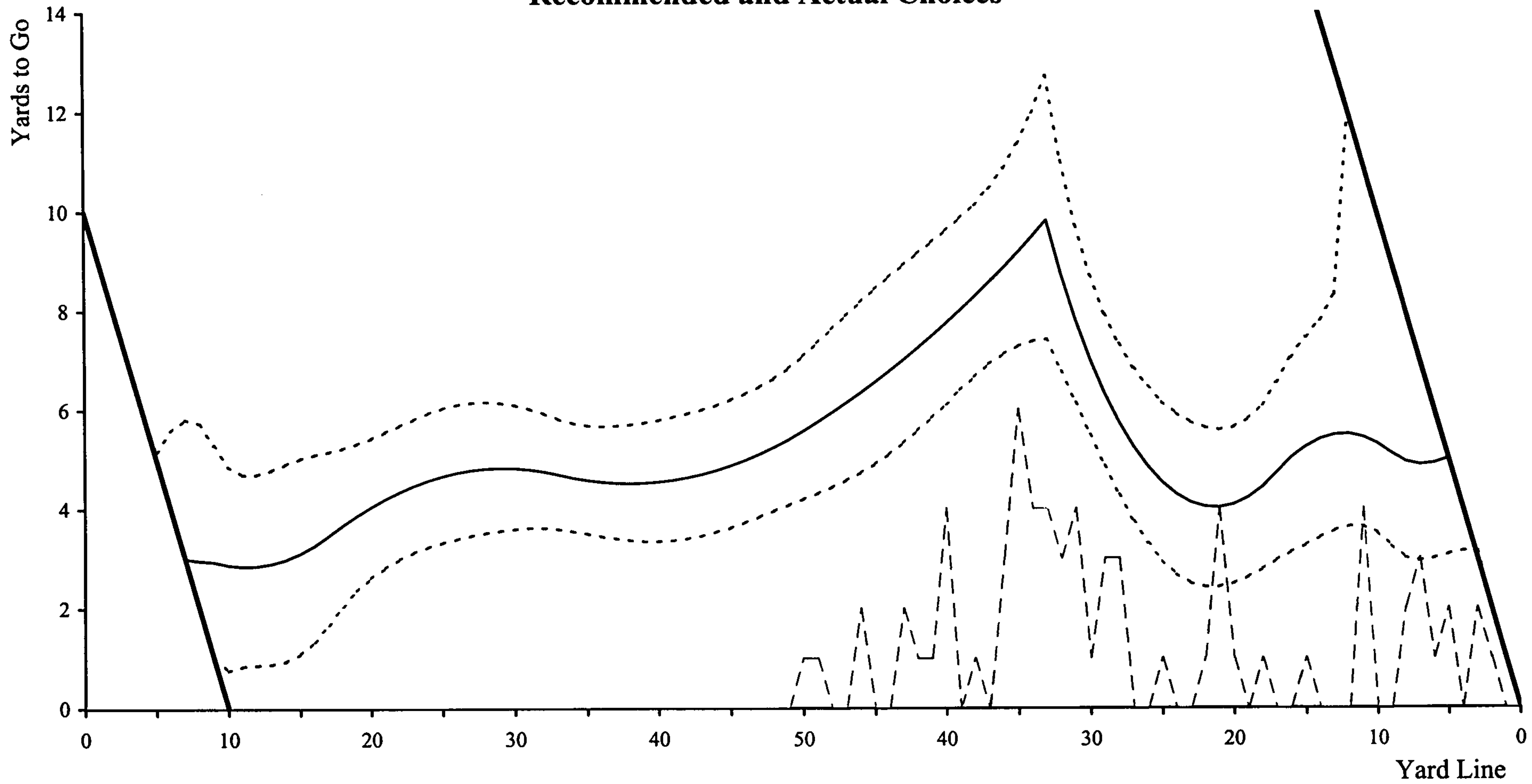
The Differences between the Values of Going for It and of Turning the Ball Over



The figure shows the estimated differences between the values of going for it and the other team having the ball on the spot at a generic yard line outside the opponent's 17 (solid line) and at the opponent's 5 (dashed line). The dotted lines show the two-standard-error bands.

Figure 5

Recommended and Actual Choices



The solid line shows the number of yards to go where the estimated values of kicking and going for it are equal. The dotted lines show the two-standard-error bands. The dashed line shows the greatest number of yards to go such that when teams have that many or fewer yards to go, they go for it at least as often as they kick.