IT'S FOURTH DOWN AND WHAT DOES THE BELLMAN EQUATION SAY?

A DYNAMIC-PROGRAMMING ANALYSIS OF FOOTBALL STRATEGY

David Romer

University of California, Berkeley

February 2003

ABSTRACT

This paper uses play-by-play accounts of virtually all regular season National Football League games for 1998-2000 to analyze teams' choices on fourth down between trying for a first down and kicking. Dynamic programming is used to estimate the values of possessing the ball at different points on the field. These estimates are combined with data on the results of kicks and conventional plays to estimate the average payoffs to kicking and going for it under different circumstances. Examination of teams' actual decisions shows systematic, overwhelmingly statistically significant, and quantitatively large departures from the decisions the dynamic-programming analysis implies are preferable.

I am indebted to Ben Allen, Ryan Edwards, Mario Lopez, Travis Reynolds, and Evan Rose for outstanding research assistance, to Christina Romer for invaluable discussions, and to numerous colleagues and correspondents for helpful comments and suggestions.

I. INTRODUCTION

A football team takes the opening kickoff and drives down the field. The drive stalls, and the team faces fourth down and goal on its opponent's 2-yard line. When the team tries for a field goal rather than a touchdown, there are some boos from the fans, but the consensus among the fans and commentators is that it has made the right decision. But has it?

Two considerations that are obviously relevant to the team's decision are the points involved (three for the field goal, seven for the touchdown) and the probabilities of success (essentially a hundred percent for the field goal, roughly forty percent for the touchdown). But another consideration is what happens after the attempt. If the team scores, it must kick off; on average this leaves the opposing team with the ball around its 27-yard line. If the team tries for a touchdown and fails, its opponent gets the ball on the spot -- that is, roughly on its 2-yard line. Thus we need to know how much better it is to leave the opponent with the ball on its 2-yard line than on its 27.

The values of these situations depend on the probabilities of the possible subsequent situations, and their values. Those values depend in turn on the probabilities of the possible subsequent situations, and their values; and so on. Thus this is a problem that is naturally analyzed using dynamic programming. To carry out this analysis, Section II uses data from over 700 regular season National Football League games in the 1998, 1999, and 2000 seasons to estimate the values of having a first down at any point in the field, as well as the value of a kickoff. To avoid the complications introduced when one team is well ahead or when the end of a half is approaching, the analysis focuses on the first quarter.

Section III uses the results of this dynamic-programming analysis to examine teams' decisions on fourth down. To estimate the value of kicking at various points on the field, I simply 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.

The results are striking. The analysis implies that teams should be quite aggressive. A team facing fourth and goal is better off on average trying for a touchdown as long as it is within 5 yards of the end zone. At midfield, being within 5 yards of a first down makes going for it on average desirable. Even on its 10-yard line -- 90 yards from a score -- a team within 3 yards of a first down is better off on average going for it. In practice, however, teams almost always kick on fourth down early in the game. The only significant exceptions occur 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, and on plays where the team has one yard to go and is near the opponent's goal line. Even in those cases, however, teams are much more conservative than the recommendations of the dynamic-programming analysis.

Section IV considers various possible complications and biases. I conclude that they do not change the basic messages of the analysis.

Section V considers the results' quantitative implications. The estimates imply that following the prescriptions of the dynamic-programming analysis on fourth-down plays in the first quarter would raise a typical team's probability of winning a game by roughly half a percentage point. Since the analysis covers only about one-thirtieth of all plays, this suggests

that the overall gains from better strategy may be quite large.

Before proceeding, there is an obvious question that needs to be addressed. To put it bluntly, why should anyone care about people in funny costumes chasing an inflated pigskin around a field? There are four reasons.

First, the analysis demonstrates the ability of mathematical, statistical, and economic tools to provide insights into a subject that at first glance appears to have nothing to do with mathematics, statistics, or economics.

Second, the behavior of professional football teams provides a powerful test of the standard assumption of complete optimization of simple objective functions. There are by now many examples where the predictions of simple models of maximization appear to be contradicted by agents' behavior. But most of these examples involve individuals rather than firms, and they usually involve low-stakes environments. Determining whether the failures carry over to high-stakes business environments is extremely difficult: data are very limited, and the decision problems are usually so complicated that determining that a decision represents a departure from profit-maximization is almost impossible. But, as Thaler (2000) stresses, sports provides an exception to this pattern. Data are copious and detailed. And for most on-field decisions, the problem of maximizing profits reduces to the simpler problem of maximizing the probability of winning: it is difficult to think of any plausible way that decisions about, say, which relief pitcher to bring in or whether to try a two-point or a three-point shot affects a team's profits other than through its impact on the probability of winning. Further, the predictions of simple models of optimization seem especially likely to hold in the case of fourth-down decisions in professional football: winning is valued enormously (as shown by the very high salaries commanded by high-quality players), the participants almost universally claim that winning is their key objective, the decisions arise repeatedly, and decisions are not just publicly observed but nationally televised.

Third, as I explain in Section VI, football teams' behavior provides evidence not just

about the existence of failures of predictions of simple models of maximization, but about the nature and sources of those failures.

Finally, many people enjoy watching and discussing football. For them, the analysis may be of interest in its own right.

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 the hypothesis that firms maximize profits -- are less common. Two recent examples are Walker and Wooders (2001), who 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 should be considerably more aggressive on fourth downs, but their exact conclusions differ substantially from mine. I discuss the particulars of how my analysis is related to these studies below.

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).[1] 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 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 left.

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 $i \neq 100$. 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

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

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 $\sum_i D^i_{gt}V_i$. Thus we have

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

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: $e_{gt} = [P_{gt} + B_{gt}\sum_i D^i_{gt+1}V_i] - E[P_{gt} + B_{gt}\sum_i D^i_{gt+1}V_i]$. 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

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

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.[2]

There is one final issue. There are 101 V_i's to estimate. Even with three years 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

---

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

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.[3]

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.[4] 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.[5]

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

---

[3] 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.

[4] The data for two games in 1999 and two games in 2000 were missing from the website.

[5] 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.

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, $\widetilde{D}^i_{gt}$ a dummy that equals 1 if the next situation after kick g,t is a situation of type i, $\widetilde{P}_{gt}$ 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 $\widetilde{B}_{gt}$ 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

$$\widetilde{P}_{gt} + \widetilde{B}_{gt}\sum_i \widetilde{D}^i_{gt} V_i = \sum_i K_i A^i_{gt} + u_{gt}. \tag{5}$$

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.[6]

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

---

[6] 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 four 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 five 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.

bands.[7] On the team's 1-yard line, the estimated value of kicking is -2.7; this is the same as the value of the other team having a first and 10 on one's 35. 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," reaching a low of 1.5 (a punt of only 25 yards) at the opponent's 33. It 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, as mentioned in the introduction, 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

---

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

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 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.[8] The resulting estimates of the difference between the values of going for it and of turning the ball over show no 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.[9]

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:

---

[8] Specifically, I divide the field at the 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 equation (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.

[9] 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.

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

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

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

At the 17, where the two functions meet, I constrain both their level and their derivative with 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.[10] 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.[11] 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

[10] 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 irrelevant to decisions about going for it from affecting the results, plays where the team had more than 20 yards to go are excluded.

[11] 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.

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.[12]

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 6.6 yards at the opponent's 45 and peaks at 9.5 on the opponent's 33. As the team gets into field-goal range, the critical value falls rapidly; its lowest point is 4.1 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.

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

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.[13] The critical values are estimated fairly precisely.

Although these findings contradict the conventional wisdom, they are quite intuitive. Consider two examples. The first 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 on average better off than its opponent if it goes for it. If the team punts, the opponent will on average end up with a first and 10 around its 15. Both standard views about football and the analysis in Section II suggest that the receiving team and the punting team 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 second example is fourth and goal on the 2. If the team goes for a touchdown, it has about a three-sevenths chance of success. Thus kicking a field goal and going for a touchdown have essentially the same expected payoff in terms of immediate points. But when the team goes for a touchdown, the next situation is on average much better for it: about four-sevenths of the time, the opponent has a first and 10 near its goal line rather than receiving a kickoff. Thus the aggressive strategy is again 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

[13] 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.

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.

<u>Actual Choices</u>. Teams' actual choices are dramatically more conservative than the choices recommended by the dynamic-programming analysis. On the 1575 fourth downs in the sample where the analysis implies that teams are on average better off kicking, they went for it only 7 times. But on the 1100 fourth downs where the analysis implies that teams are on average better off going for it, they kicked 992 times.[14]

The dashed line in Figure 5 provides a simple summary of 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 equal is typically rejected with a t-statistic between 3 and 7.[15]

---

[14] These figures exclude the 22 cases where we cannot observe the team's intent because of a penalty before the snap.

[15] 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. 3 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

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 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, teams are conservative even in situations where they should be risk loving over points scored. Consider teams that are trailing by at least 4 points in the third quarter. Since these teams need more than a field goal even to tie the game, the same logic that implies that teams that are ahead or tied should be risk averse implies that these teams should be risk loving. But they are only slightly more likely than teams in the first quarter to go for it on fourth down: in the 363 such cases in the sample where the dynamic-programming analysis implies that going for it is on average preferable if teams are risk neutral, the teams

---

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.

kicked 270 times.

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 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_i'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_i'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.

Additional Information. When a team is choosing between kicking and going for it, it 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 blindly follow the recommendations of the dynamic-programming analysis.

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 532 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 8 times. Similarly, on the 183 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. This prediction is contradicted by the data: the average difference between the outcome of plays where teams go for it on fourth down and what one would expect based on third-down plays is a substantively and statistically small negative number.

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. But 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 reserve.

Thus, the direction of the bias from using third-down plays is likely 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-down plays 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-down plays 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-down plays 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.

Momentum. Failing on fourth down could be costly not just in terms of possession and field position, but in terms of energy and emotions. As a result, it might be more costly for the other team to have the ball at a given point as a result of stopping a fourth-down attempt than for it to have the ball there 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_i'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.[16]

Partial Equilibrium versus General Equilibrium. The analysis looks at decisions on individual fourth downs taking all other decisions as given. But there are at least two ways that these decisions could affect other aspects of teams' choices.

First, making different decisions on fourth downs will alter the V's. By being more aggressive on fourth downs, the offense can do better. This suggests that the payoff to obtaining a first down is greater than the estimates imply, and that the importance of field position may be smaller. Thus, the analysis may understate the value of going for a first down relative to punting, and may overstate the value of going for a touchdown relative to attempting a field goal. 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 they suggest that the analysis understates the gap between actual and optimal strategies.

Second, the knowledge that the offense will be more aggressive on fourth downs can affect behavior on other downs. Most importantly, greater aggressiveness 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 the 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

[16] Looking at the much smaller sample of 115 fourth-down attempts produces qualitatively similar but much less precisely estimated results. The point estimates again go slightly against the predictions of the momentum hypothesis, and are again statistically and substantively insignificant.

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, this effect is unlikely to lead to substantial overestimates of the value of improved fourth-down decisions.

V. QUANTITATIVE IMPLICATIONS

An obvious question is whether the gains to better strategy are important. To estimate the potential benefits of going for it more often on fourth downs, I first find all cases in the sample where a team kicked when the difference between the estimated values of going for it and kicking was positive. There are 992 such cases in the 732 first quarters in the sample, or an average of 0.68 cases per team-quarter. I then average the estimated values of the expected gain from going for it in these cases. The average is 0.37 points. Thus the expected payoff to a typical team of being more aggressive on fourth downs in the first quarter is approximately 0.25 points per game. The evidence presented in the previous section suggests that an additional point raises the probability of winning by around 2.3 percentage points. Thus, it appears that better decisions about whether to go for it on fourth downs in the first quarter could raise a team's probability of winning a typical game by about 0.6 percentage points.

Fourth downs in the first quarter are only about one-thirtieth of all plays. The tools used in this paper, extended to account for time and score, could be applied to decisions about going for it on fourth downs in other quarters. And the evidence presented in Section IV about fourth-down choices in the third quarter by trailing teams suggests that there may

be significant benefits from better fourth-down decisions outside the first quarter. These teams are only slightly more likely to go for it than teams in the first quarter. But these teams' strategies should be even more aggressive than the strategies that the analysis implies they should follow in the first quarter.

In addition, tools like those used here could be used to analyze other decisions. Some obvious examples are punting versus field-goal attempts, one-point versus two-point extra-point attempts, regular versus onside kickoffs, accepting versus declining penalties, and some aspects of clock management. Similar tools might also be useful for analyzing choices among broad categories of plays; for example, one could ask if teams do better on average when they attempt running or passing plays on first down.

These considerations suggest substantial uncertainty about the overall potential benefits from better strategy. At one extreme, a figure of 2.3 percentage points -- four times the estimate based on fourth downs in the first quarter -- is almost certainly too low. At the other extreme, there is no evidence that the potential benefits on an average play are as high as the average benefit on fourth-down plays in the first quarter. Thus, a figure of 18 percentage points -- 30 times the estimate from first-quarter fourth downs -- is almost certainly too high.

Even estimates toward the low end of this range would be quantitatively important. For example, suppose that the overall gain from better choices is an increase of 4.7 percentage points in the probability of winning an average game. This corresponds to the case where the gains from better fourth-down choices in each of the other quarters are as large as in the first quarter, and where half of the total potential benefits from improved strategy come entirely from the decision between kicking and going for it on fourth downs. Thus this estimate appears conservative. Yet it implies that better strategy would allow a typical team to win one more game in three seasons out of four.

VI. CONCLUSION

The behavior of National Football League teams on fourth down appears to depart systematically from the behavior that would maximize their chances of winning. The departure is quantitatively important and overwhelmingly statistically significant. And it cannot be explained by rational risk aversion, information known to teams that is omitted from the analysis, or various other complications.

The departure from maximizing behavior is 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.

Two other forces could, however, lead to systematically conservative choices. 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 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).

It seems unlikely that such preferences can fully explain the results. First, most fans and players appear to want their teams to go for it more often. Second, and more importantly, the amount of risk aversion over probabilities needed to account for the results is very large. For example, consider a team facing fourth and goal on the 1-yard line early in the game. The estimates suggest that if it attempts a field goal, its chances of winning the game are about 55.5 percent. If it goes for a touchdown, it has about a 45 percent chance of success, in which case its chances of winning are about 65 percent, and a 55 percent chance of failure, in which case its chances of winning are about 54 percent; thus the overall chance of winning is about 59 percent.[17] In other words, by attempting the field goal 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 four times as much weight on reductions in the probability of winning below what can be attained by the field-goal attempt as on increases in the probability of winning from that level.

The other 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

---

[17] These figures are computed by combining the estimated V's with the estimate in Section IV that each point raises the probability of winning by 2.3 percentage points.

context than is appropriate for maximizing their chances of winning. As with risk aversion over probabilities, such imperfect optimization would have to be substantial to account for the results.[18]

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

---

[18] The departures from win-maximizing choices need not stem entirely from forces promoting conservatism. There are at least three other factors that 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. 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. And third, failed fourth-down attempts may be more memorable or salient than successful ones, so that decision-makers underestimate the chances of success on fourth down.

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.

outcomes), subjects exhibit considerable risk aversion over probabilities.[19]

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 act risk averse 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-after decision-makers in the industry, faced with decisions that arise repeatedly, make choices that differ systematically and in a quantitatively important 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 discussed in Section III) 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,

---

[19] 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.

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.

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.

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.

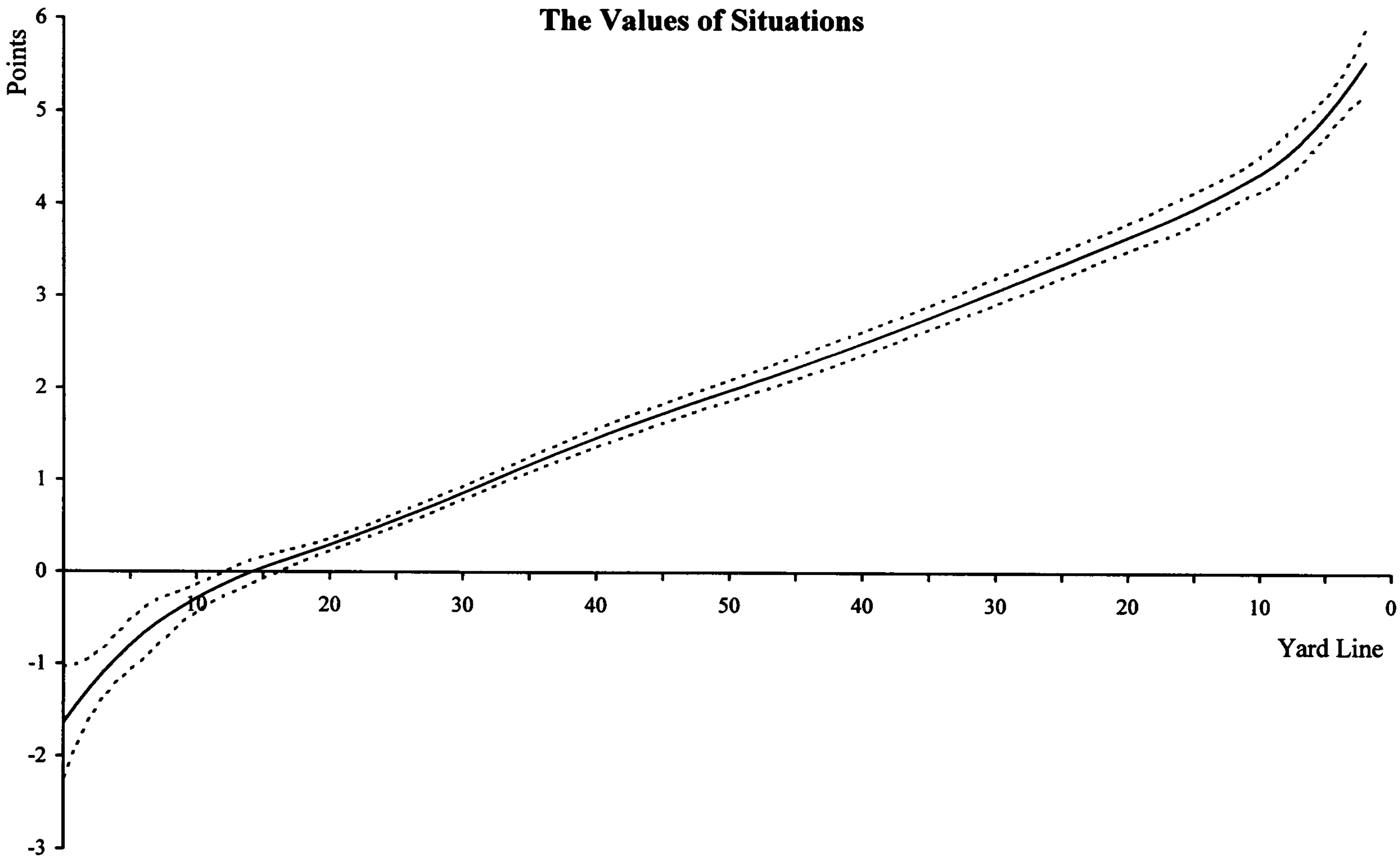
Scharfstein, David S., and Jeremy C. Stein. 1990. "Herd Behavior and Investment." American Economic Review 80 (June): 465-479.

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.

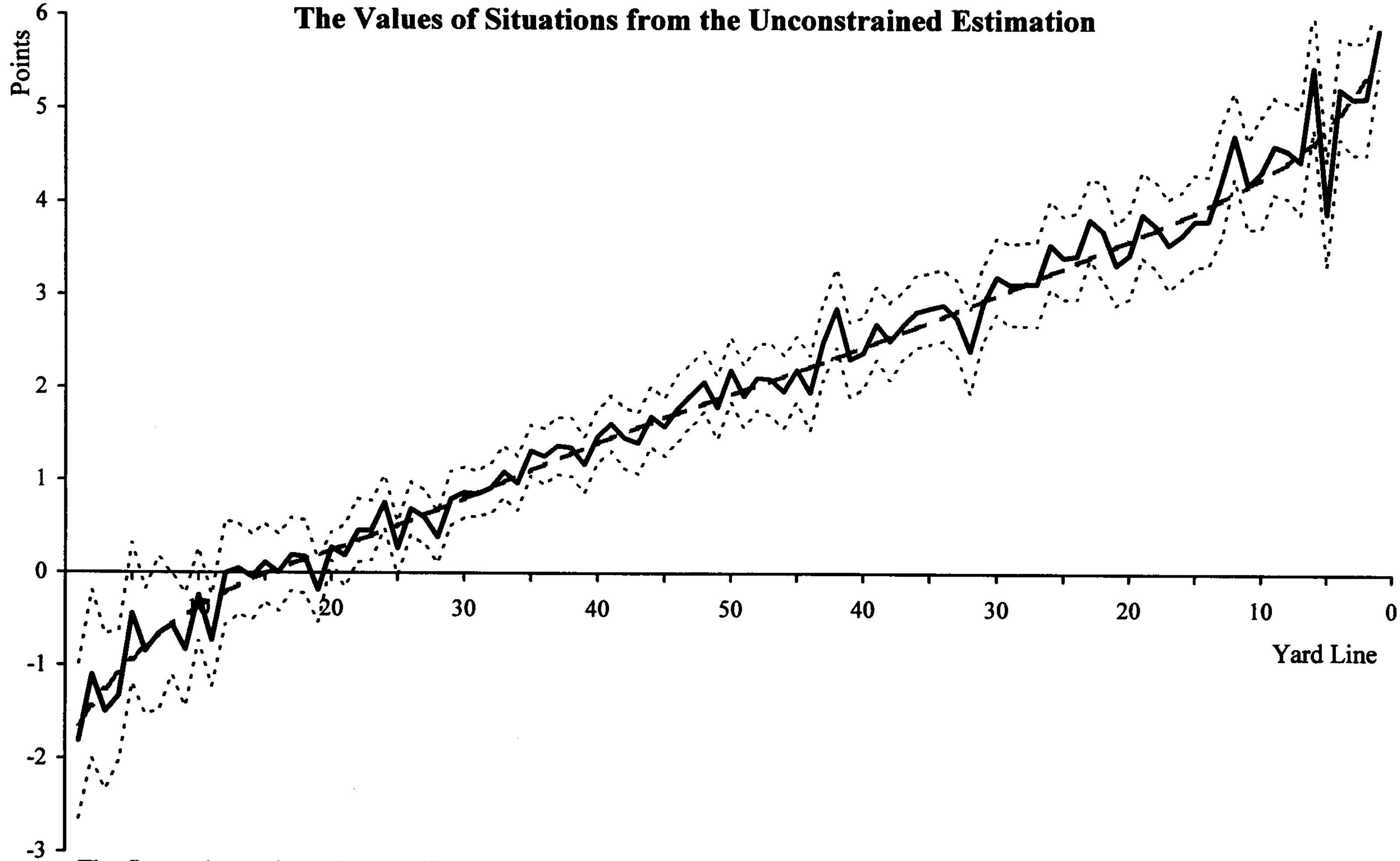
Figure 1



The 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

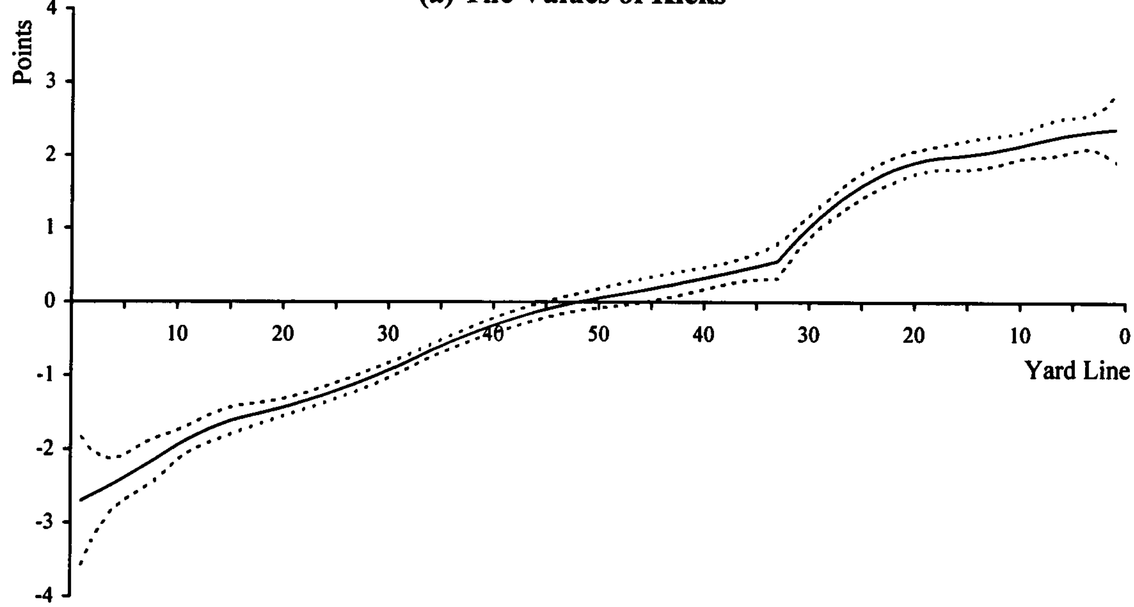
The Values of Situations from the Unconstrained Estimation



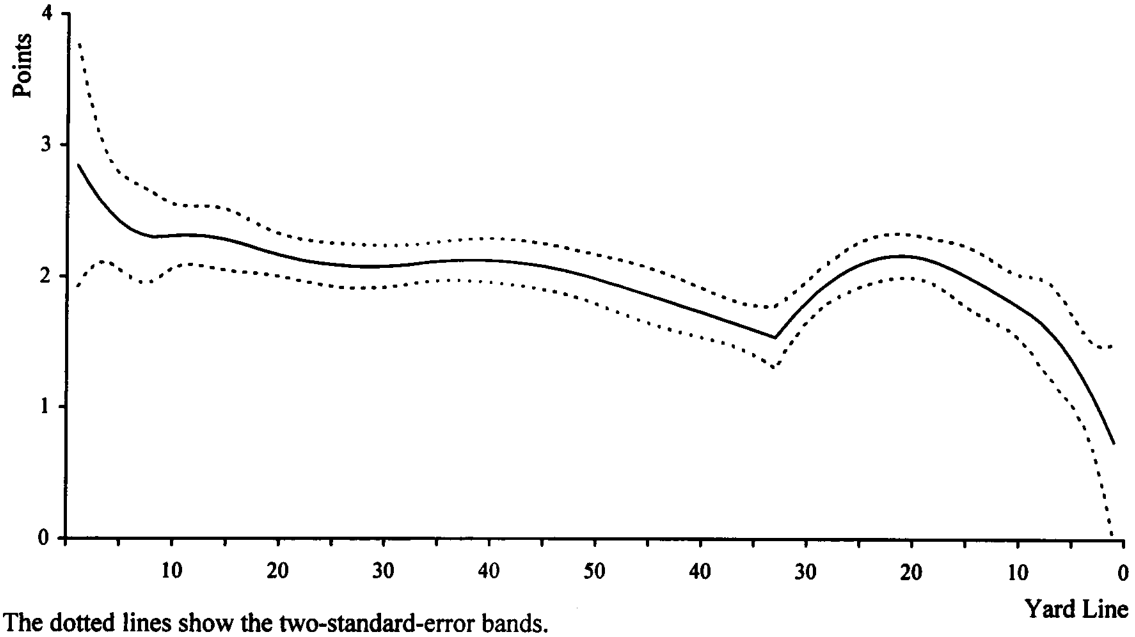
The figure shows the estimates of the V's from the unconstrained estimation. The dotted lines show the two-standard-error bands. The dashed line shows the estimates of the V's from the splined estimation.

Figure 3

(a) The Values of Kicks



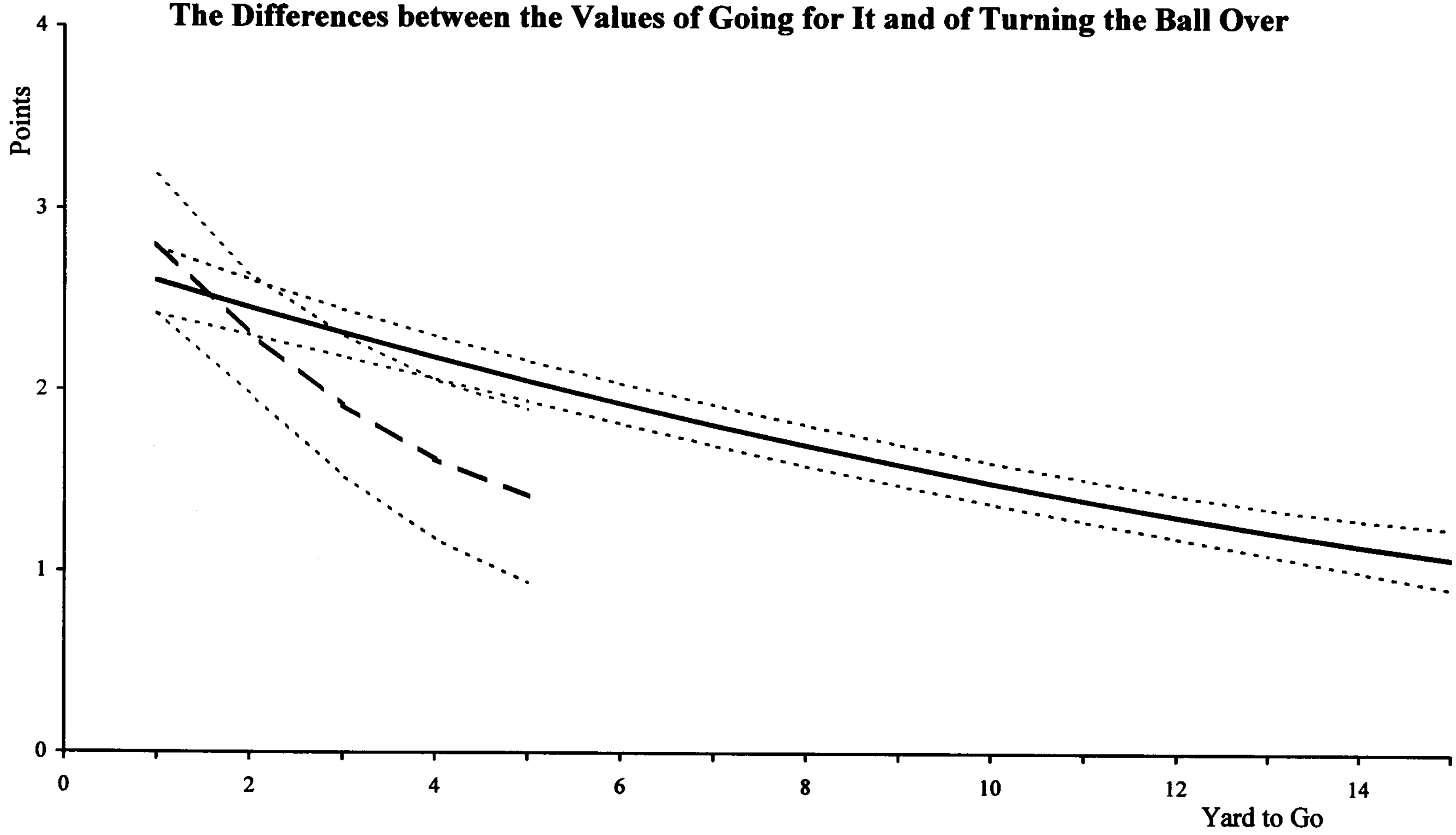
(b) The Differences between the Values of Kicks and of Turning the Ball Over



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

Figure 4

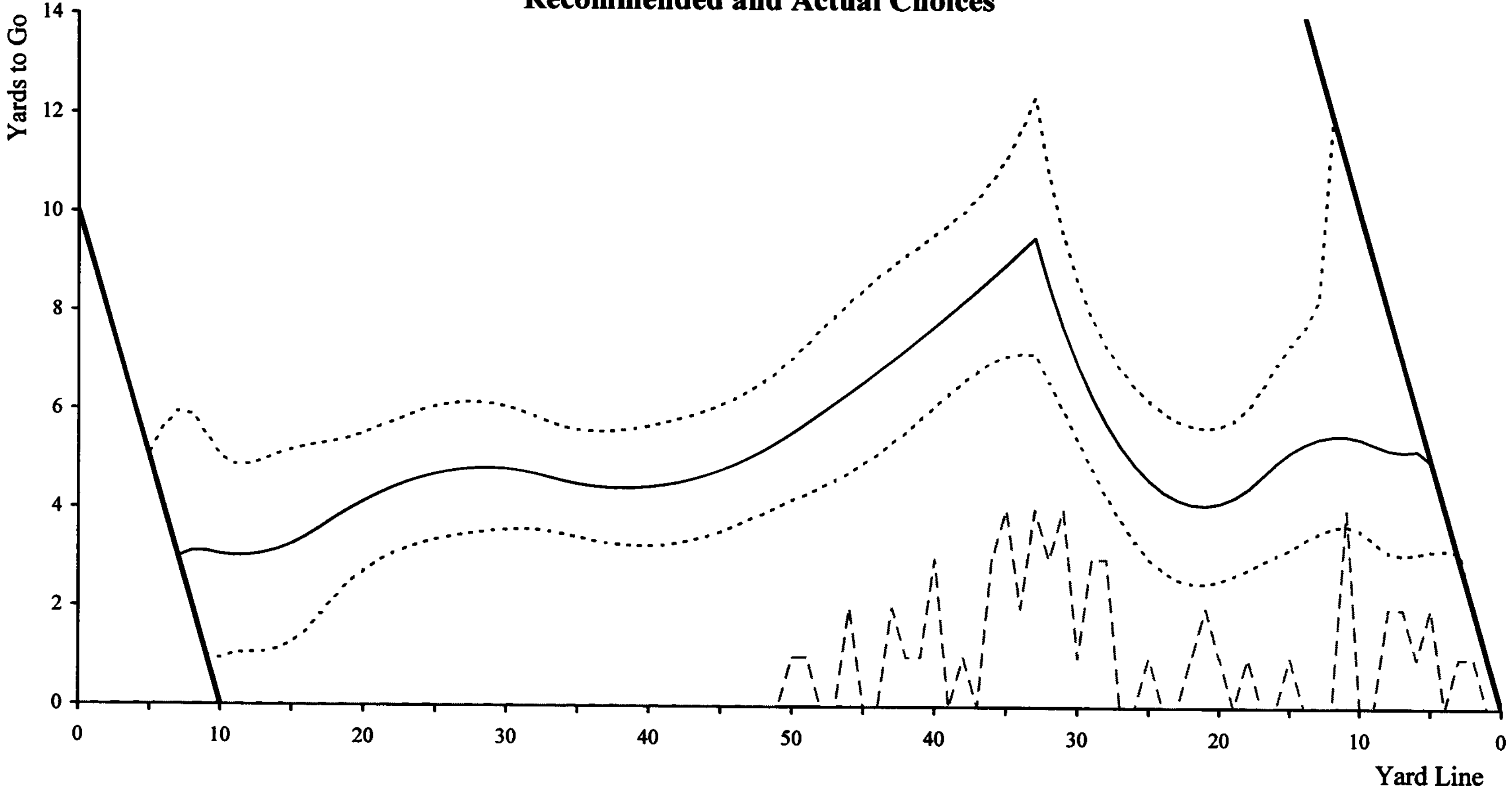
The Differences between the Values of Going for It and of Turning the Ball Over



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.