

NBER WORKING PAPER SERIES

IT'S FOURTH DOWN AND WHAT DOES THE BELLMAN EQUATION SAY?
A DYNAMIC-PROGRAMMING ANALYSIS OF FOOTBALL STRATEGY

David Romer

Working Paper 9024
<http://www.nber.org/papers/w9024>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2002

I am indebted to Ben Allen, Ryan Edwards, Mario Lopez, Travis Reynolds, and Evan Rose for outstanding research assistance, and to Christina Romer for invaluable discussions. The views expressed herein are those of the author and not necessarily those of the National Bureau of Economic Research.

© 2002 by David Romer. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

It's Fourth Down and What Does the Bellman Equation Say?
A Dynamic-Programming Analysis of Football Strategy
David Romer
NBER Working Paper No. 9024
June 2002
JEL No. L10, D21, L83

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.

David Romer
Department of Economics
549 Evans Hall
University of California
Berkeley, CA 94720-3880
and NBER
dromer@econ.berkeley.edu

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 a few scattered 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.

Finally, Section VI offers concluding remarks.

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 numerous examples of cases where the predictions of simple models of maximization appear to be contradicted by agents' behavior. In the case of professional football, the usual assumption of profit-maximization implies that in their on-field behavior, teams should act to maximize their probabilities of winning. This is a prediction that appears especially likely to hold: the agents involved are firms rather than households, the rewards to winning are enormous, the participants almost universally claim that winning is their key objective, and decisions are not just publicly observed but nationally televised.

Third, as I explain in the conclusion, football teams' behavior can help us distinguish between competing explanations of failures of predictions of simple models of maximization.

Finally, many people enjoy watching and discussing football. For these individuals, the analysis may be of interest in its own right.¹

¹ Many papers use sports data to test economic theories. Examples include Gwartney and Haworth (1974), McCormick and Tollison (1984), Ehrenberg and Bognanno (1990), Nardinelli and Simon (1990), and Fort and Quirk (1995). Attempts to use economic tools to analyze sports strategy are less common. Two examples are Chiappori, Levitt, and Groseclose (2000), who consider strategies on penalty kicks in soccer, and Walker and Wooders (2001), who examine serving strategies in tennis. In contrast to this paper,

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).² 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 does not stop at the end of the first quarter, 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.

Chiappori, Levitt, and Groseclose find no evidence of departures from optimal strategies, and Walker and Wooders find only minor deviations.

² The appendix summarizes the rules of football that are relevant to the paper.

Let g index games and t index situations within a game. Let D_{gt}^i 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, $D_{gt}^{100} = 1$ for all g and $D_{gt}^i = 0$ 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 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 's that equals one in that situation, times the V_i for that situation, times B_{gt} . That is, it equals $B_{gt} \sum_i D_{gt+1}^i V_i$.

The value of situation g, t as of that situation to the team with the ball 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 $\sum_i D_{gt}^i V_i$. Thus we have

$$\sum_i D_{gt}^i V_i = E[P_{gt} + B_{gt} \sum_i D_{gt+1}^i V_i], \quad (1)$$

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 :
 $e_{gt} = [P_{gt} + B_{gt} \sum_i D_{gt+1}^i V_i] - E[P_{gt} + B_{gt} \sum_i D_{gt+1}^i V_i]$. Using this definition, we can rewrite (1) as

$$\sum_i D_{gt}^i V_i = P_{gt} + B_{gt} \sum_i D_{gt+1}^i V_i - e_{gt}, \quad (2)$$

or

$$\begin{aligned} P_{gt} &= \sum_i D_{gt}^i V_i - B_{gt} \sum_i D_{gt+1}^i V_i + e_{gt} \\ &= \sum_i V_i (D_{gt}^i - B_{gt} D_{gt+1}^i) + e_{gt}. \end{aligned} \quad (3)$$

To think about the estimation of the V_i 's in terms of a regression, define $X_{gt}^i = D_{gt}^i - B_{gt} D_{gt+1}^i$. Then (3) becomes

$$P_{gt} = \sum_i V_i X_{gt}^i + e_{gt}. \quad (4)$$

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 $-B_{gt} D_{gt+1}^i$ terms of the X_{gt}^i 's. Recall, however, that e_{gt} is 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 . Thus the expectation of e_{gt} conditional on knowing the situation at time g,t (that is, conditional on the D_{gt}^i 's) is zero. That is, the D_{gt}^i 's are legitimate instruments for the X_{gt}^i 's. Further, since they enter into the X_{gt}^i 's, they are almost surely correlated with them. Thus we can estimate (4) by instrumental variables, using the D_{gt}^i 's as

the instruments.³

There is one final issue. There are 101 V_i 's to be estimated. Even with three years of data, the estimates of the individual V_i 's will be somewhat 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.

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

³ There is another way of describing how the V_i 's are estimated. 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 these assumed V_i 's and the actual P_{gt} 's. Repeat the process using these new V_i 's as an input, and iterate until the process converges. One can show that this procedure yields estimates identical to those produced by the instrumental-variables approach.

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

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 roughly the 15. That is, the estimates imply that a team should be indifferent between a first and 10 on its 15-yard line 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_i '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_i '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.

the latter is conceptually different from the former, I present evidence in Section IV that it is an excellent approximation.

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 the 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_{gt}^i be a dummy that equals 1 if kick g,t is from yard line i , \tilde{D}_{gt}^i 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

$$\tilde{P}_{gt} + \tilde{B}_{gt} \sum_i \tilde{D}_{gt}^i V_i = \sum_i K_i A_{gt}^i + u_{gt}. \quad (5)$$

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_{gt}^i '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

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

Figure 3 reports the results. Panel (a) shows the estimated value of kicking as a function of the team's position, together with the two-standard-error 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

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 the team needs for a first down.

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.

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 had been fourth-down plays, and then 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, the difference between the values of going for it and of turning the ball over 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. 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. But 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 As a result of the large number of parameters that must be estimated with this approach, the values of going for it in each circumstance are estimated quite imprecisely. Since estimating the difference between the values of going for it and turning the ball over provides a convenient way of imposing plausible restrictions, I focus on this difference.

Specifically, I proceed as follows. Let G_{iy} denote the value of going for it on yard

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.

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 :

$$G_{iy} + V_{i'} = a_0 + a_1y + a_2y^2. \quad (6)$$

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

$$G_{iy} = b_0 + b_1y + b_2i + b_3y^2 + b_4yi + b_5i^2 + b_6y^2i + b_7yi^2 + b_8y^2i^2. \quad (7)$$

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 estimate 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 dotted 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

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.

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 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 dashed 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 less data 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

¹¹ This translation of average outcomes into point values, and the analogous ones in the remainder 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.

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.

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 field. 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 for a first down. 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. The analysis implies that once a team reaches its opponent's 5, it is always better off on average going for it on fourth down. The two dashed 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 a first down. 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

¹³ For example, the lower dashed line shows the point where the difference between the estimated values of going for it and kicking is twice its standard error.

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 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 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 dotted 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 are more likely to kick than to go for it. 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.

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

IV. COMPLICATIONS

This section discusses five factors that have been omitted from the analysis and that have the potential to bias the results. The first three concern the estimation of the V 's, and the final two concern the choice between kicking and going for it given the V 's.

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 very 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. Yet, as we have seen, teams are too 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 these teams 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 kicked 270 times.

Third, it is possible to obtain some direct evidence about the impacts of different numbers 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 equals 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 more 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 exactly 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 in what direction this effect 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 the effects of any selection bias.

The change in the sample has only a small effect 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 estimates provide no evidence that selection bias is important.

Partial Equilibrium versus General Equilibrium. The estimates of the V 's take teams' choices as given. But the analysis implies that teams can do better by being more aggressive. 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 implies 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.

Additional Information. When a team is choosing between kicking and going for it, it can use more information than the averages employed by 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 more. Thus it would be suboptimal for it to blindly follow the recommendations of the dynamic-programming analysis.

Additional information cannot, however, account for the large systematic departures from the implications of the dynamic-programming analysis. Over wide ranges, teams almost always kick in circumstances where the analysis implies that they would 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 that 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 on fourth down, the results will be dramatically 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 in fact 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 cost of failing is higher on fourth down than on third down. This suggests that if teams did go for it

on fourth down, they would orient their plays more toward making the yards needed for a first down (as opposed to maximizing the expected yards gained) than they do on third down. The same reasoning suggests that defenses would orient their choices more toward preventing a first down.

Two considerations suggest that the use of third-down plays to estimate what the outcomes of fourth-down plays would be is not the source of the results. First, it seems likely that the offense has more scope than the defense to affect the distribution of outcomes. If this is correct, using the outcomes of third-down plays leads to underestimates of the value of going for it. Second, the relative payoffs to different outcomes do not differ greatly between 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.

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 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, and accepting versus declining penalties. 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, however. 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 downs appears to be systematically more conservative than the behavior that would maximize their chances of winning. The departure from maximizing behavior 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 a variety of other complications.

The rewards to winning in the National Football League are enormous. Why then do teams appear not to maximize their chances of victory? One possibility, of course, is that they do. That is, there could be some important consideration that the analysis of this paper omits that implies that teams' choices are not in fact systematically too conservative. If we leave this possibility aside, there are two others. One is that the relevant actors do not want to maximize their teams' probabilities of winning. In particular, the costs of losing as a result of a failed gamble could be greater than the costs of losing from playing it safe, so that

decision-makers choose to depart from the behavior that is best in terms of the chances of winning. The additional costs could come from outside the team (in the form of increased criticism and reduced support from fans), from principal-agent considerations within the team (in the form of adverse consequences for those making the strategic decisions from individuals higher in the organization), or even from internal feelings (in the form of greater regret).

The other possibility is that the decision-makers are trying to maximize their teams' chances of winning, but that they 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 it may be that the individuals involved want to make the decisions that maximize their teams' chances of winning, but that they rely on experience and intuition rather than formal analysis, and that these lead them to depart significantly from the optimal choices.

These two possibilities have different implications about the future evolution of football strategy. If coaches follow conservative strategies because they do not want to maximize their chances of winning, their behavior will not change without changes in the forces that give rise to those objectives (such as fans' preferences or teams' structures). If coaches follow such strategies because they are optimizing imperfectly, 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 towards 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 wherever 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.

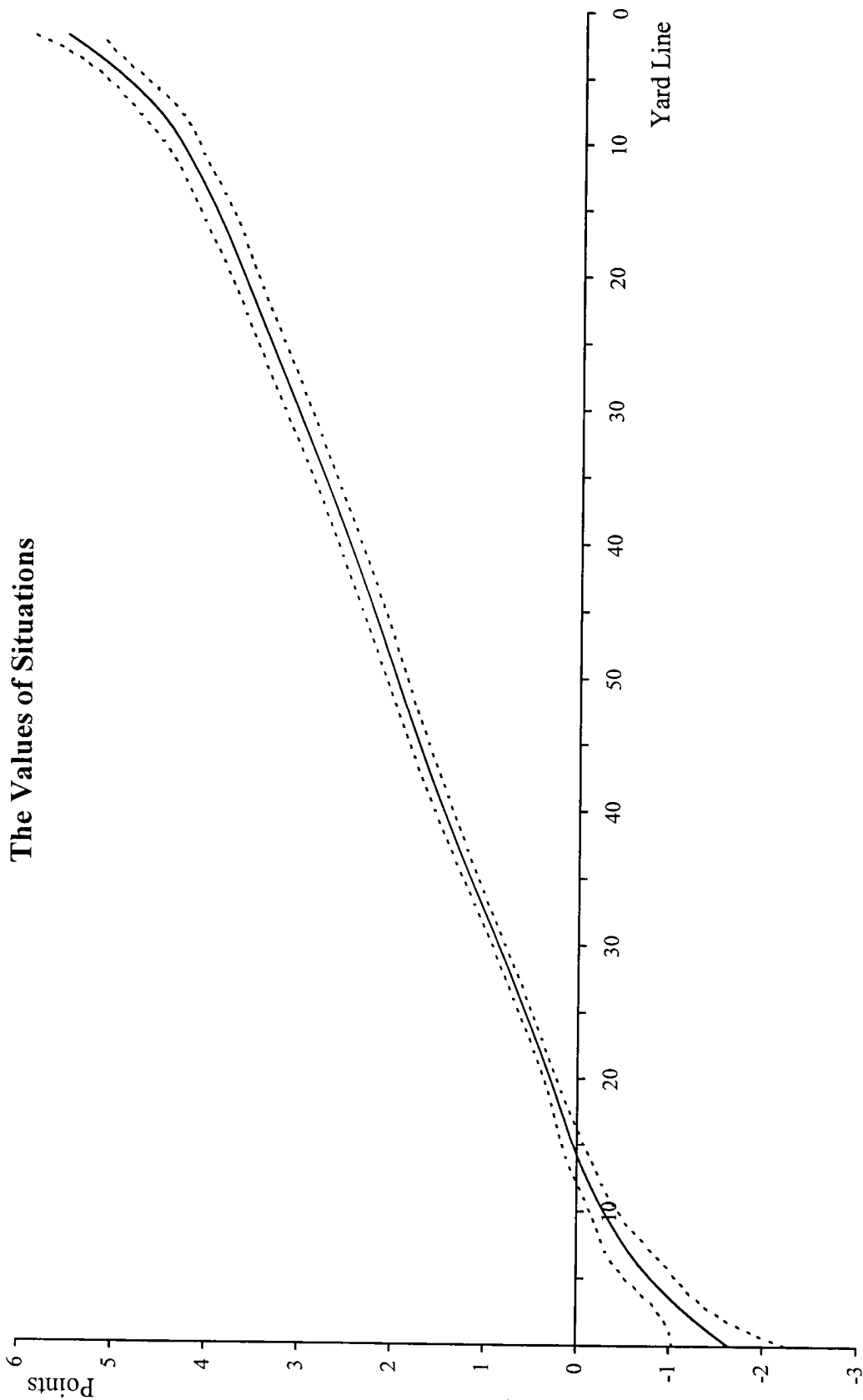
The final (and by far the least common) way to score is a "safety": if the offense is pushed back across its own goal line, the defense scores 2 points, and the offense puts the ball in play by kicking to the other team from its 20-yard line (a "free kick").

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

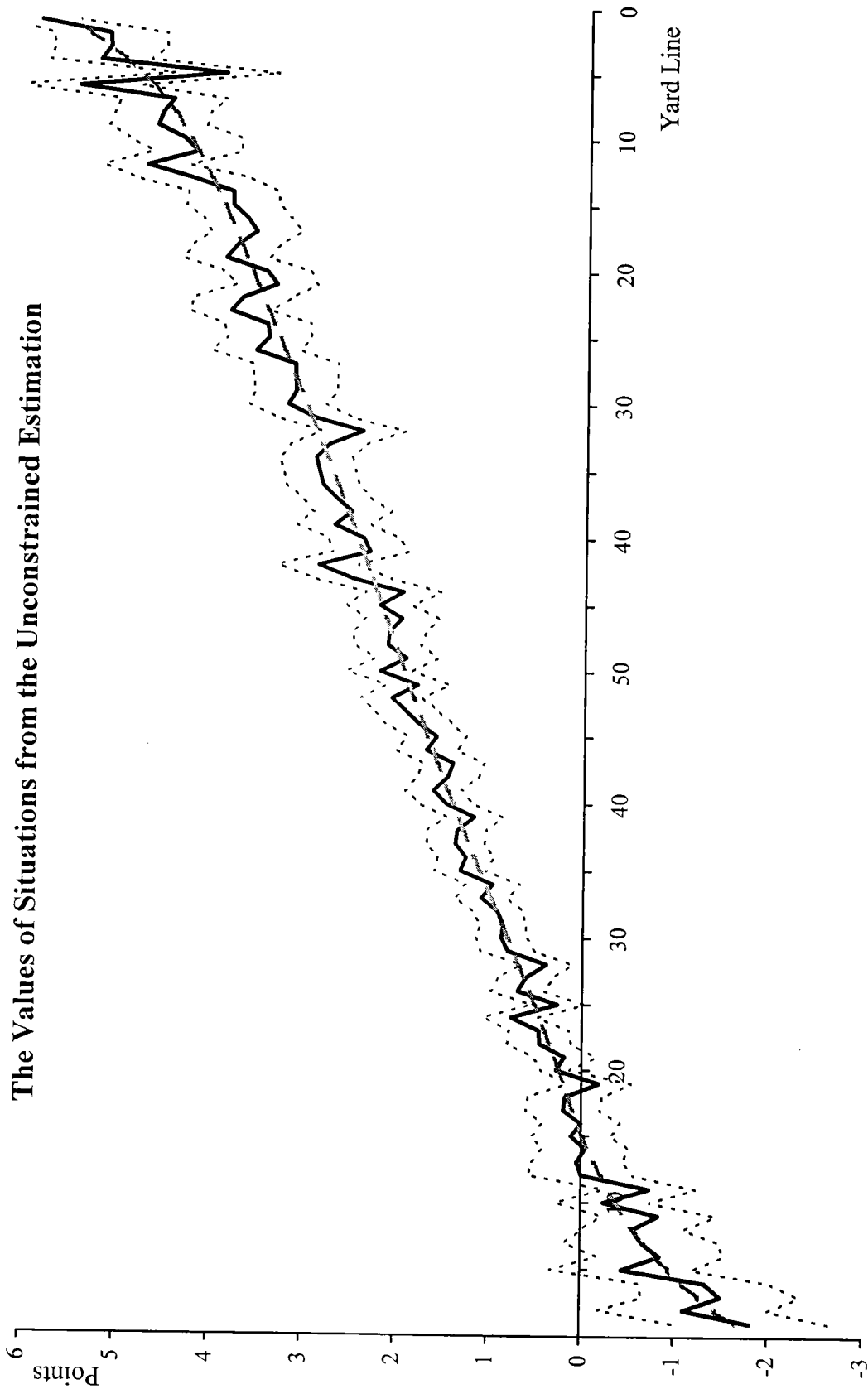
- Chiappori, P. A., S. Levitt, and T. Groseclose. 2000. "Testing Mixed Strategy Equilibria When Players Are Heterogeneous." Unpublished paper, University of Chicago (October). American Economic Review, forthcoming.
- Ehrenberg, Ronald G., and Michael L. Bognanno. 1990. "Do Tournaments Have Incentive Effects?" Journal of Political Economy 98 (December): 1307-1324.
- Fort, Rodney, and James Quirk. 1995. "Cross-subsidization, Incentives, and Outcomes in Professional Team Sports Leagues." Journal of Economic Literature 33 (September): 1265-1299.
- Gwartney, James, and Charles Haworth. 1974. "Employer Costs and Discrimination: The Case of Baseball." Journal of Political Economy 82 (July/August): 873-881.
- McCormick, Robert E., and Robert D. Tollison. 1984. "Crime on the Court." Journal of Political Economy 92 (April): 223-235.
- Nardinelli, Clark, and Curtis Simon. 1990. "Customer Racial Discrimination in the Market for Memorabilia: The Case of Baseball." Quarterly Journal of Economics 105 (August): 575-595.
- 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).

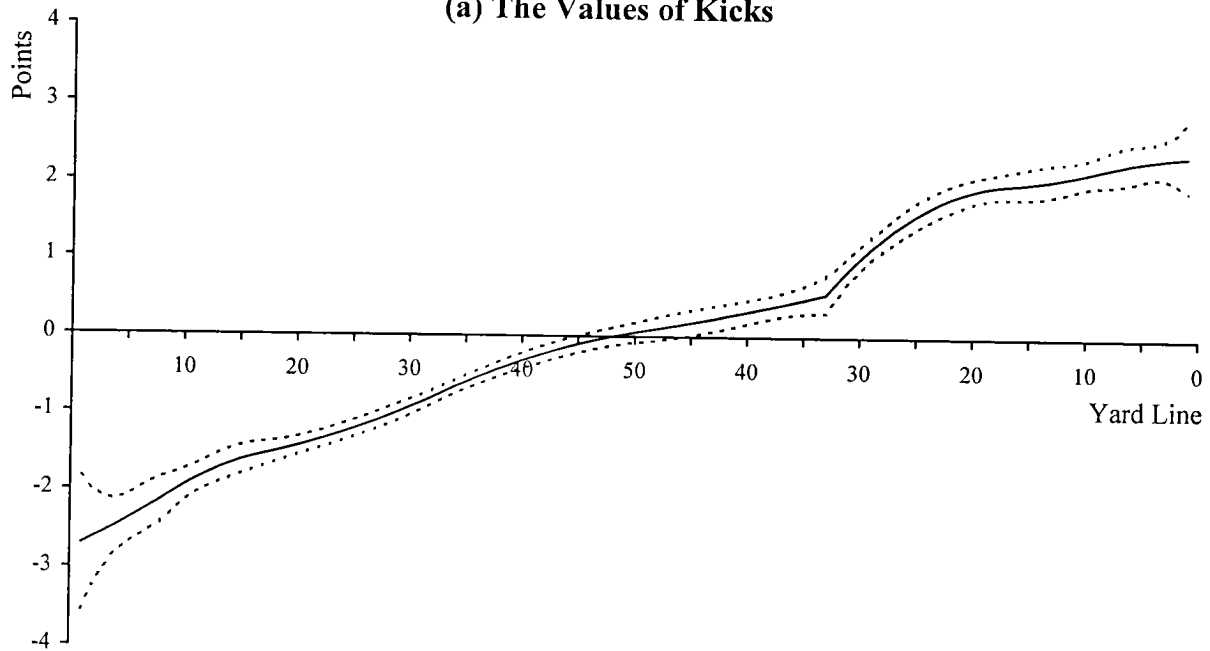
Figure 2



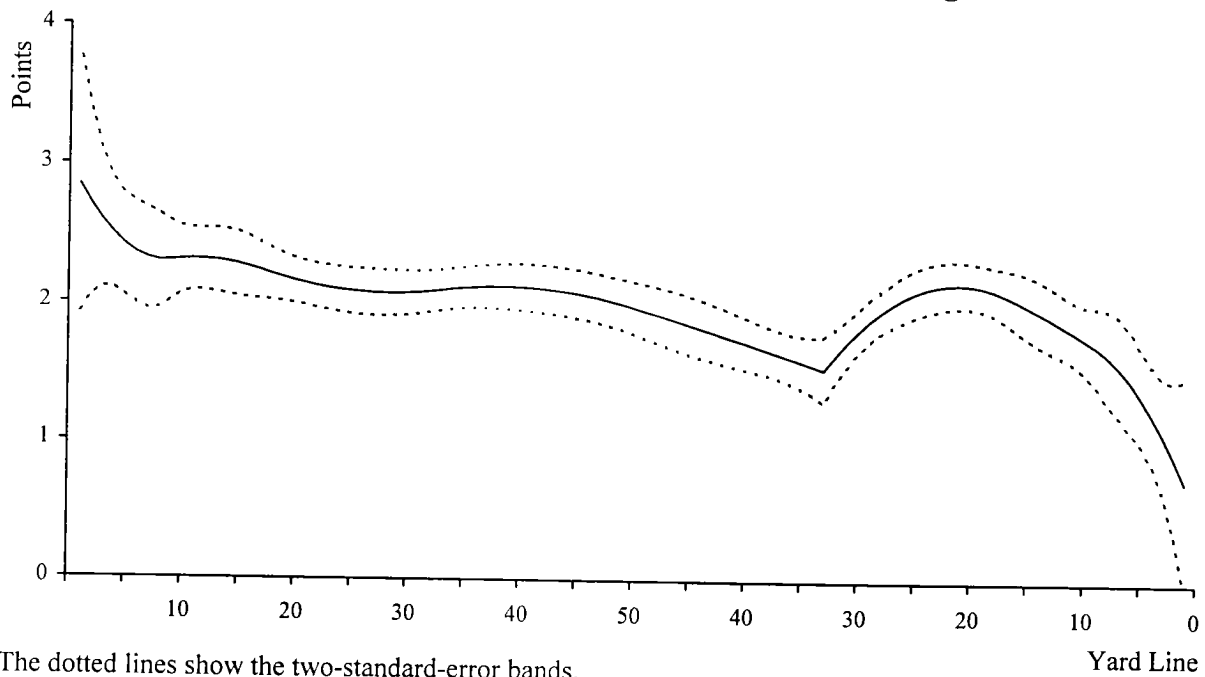
The figure shows the estimates of the V 's from the unconstrained estimation. The dotted lines show the two-standard-error bands. The dashed line shows the estimates of the V 's from the splined estimation.

Figure 3

(a) The Values of Kicks

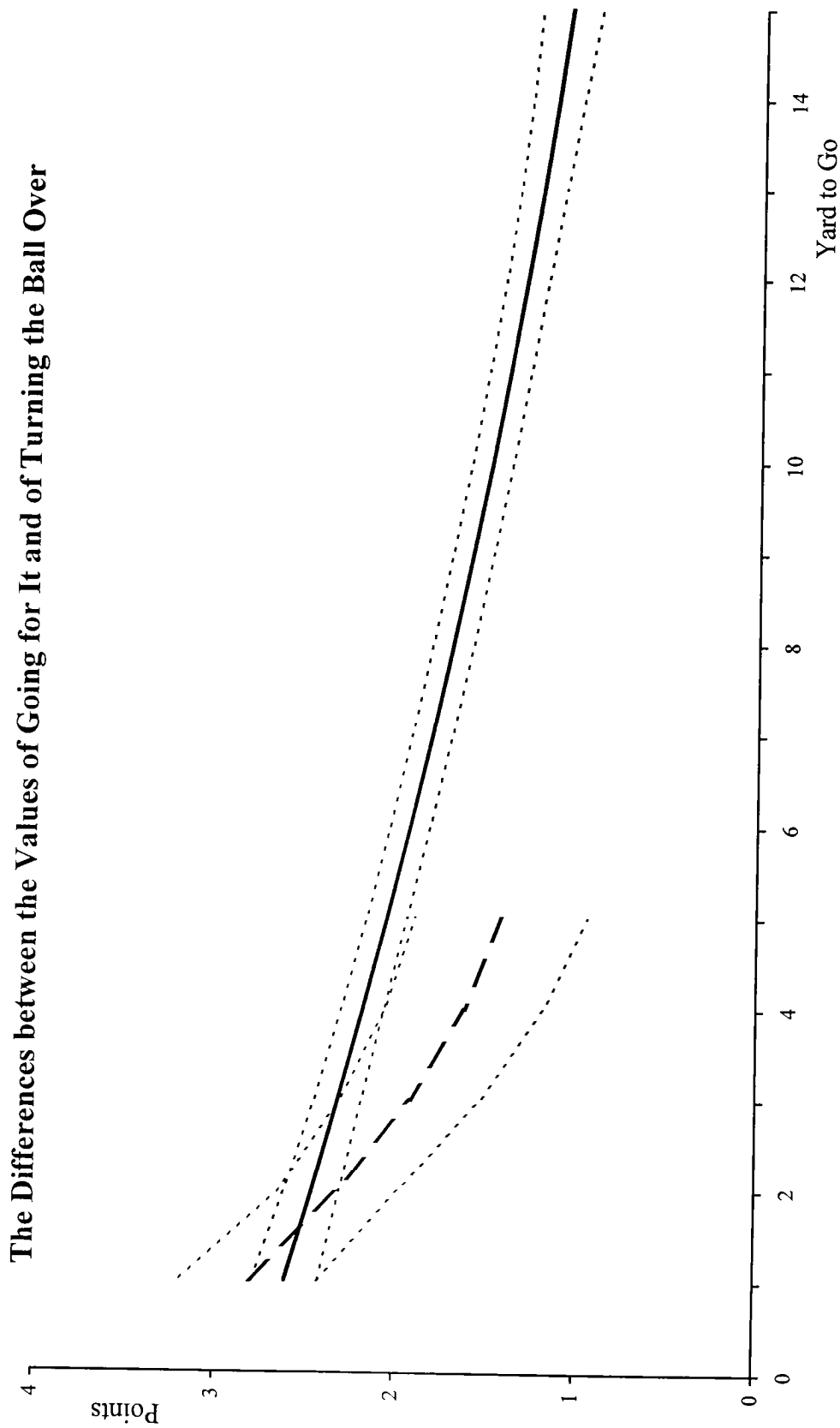


(b) The Differences between the Values of Kicks and of Turning the Ball Over



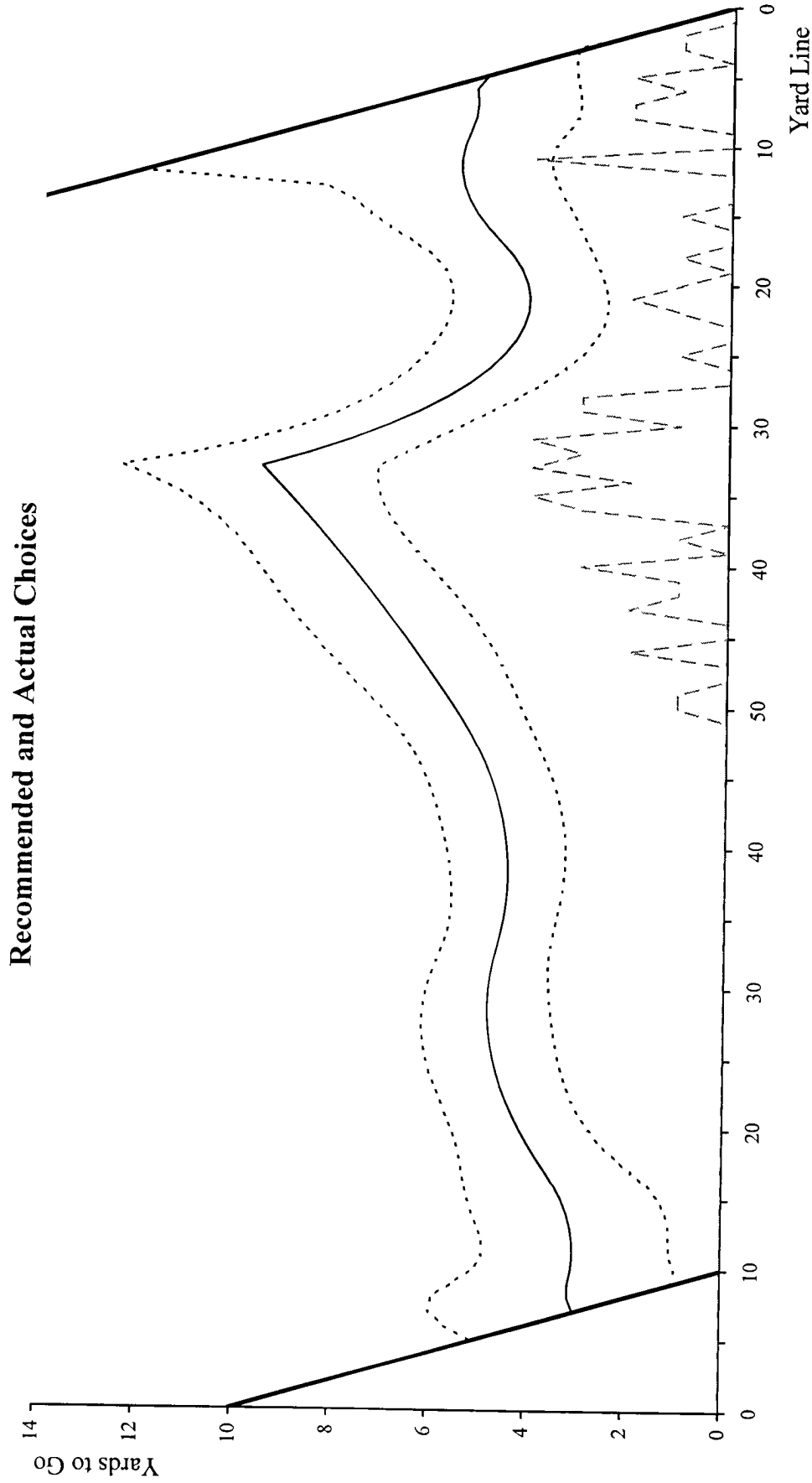
The dotted lines show the two-standard-error bands.

Figure 4



The 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



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.