D. Romer
July 2006

RESEARCHASSISTANTS.PDF

These are an edited version of the notes and instructions I wrote to the research assistants working on "Do Firms Maximize? Evidence from Professional Football." They may help to make the specifics of what was done clearer.

I have not reconciled these notes with the final version of the paper. Thus, for example, references to equation numbers may not correspond to the final version.

TABLE OF CONTENTS

DATA - A. Data on situations (first quarter)
DATA - B. Data on fourth downs (all four quarters)
DATA - C. Data on third downs (first quarter)
DATA - D. Miscellaneous data:
- -- Information about each game, including the point spread and the halftime score
- -- Data on extra-point attempts
- -- Data on coaches' salaries and turnover

ANALYSIS - A. Estimating the values of situations
ANALYSIS - B. Estimating the value of kicking
ANALYSIS - C. Estimating the value of going for it
ANALYSIS - D. Estimating the number of yards to go such that the average payoffs to kicking and going for it are equal; teams actual first-quarter decisions; potential gains from better decisions
ANALYSIS - E. Estimating the effect of points on winning
ANALYSIS - F. Analysis of teams' actual decisions in quarters 2, 3, and 4
ANALYSIS - G. Examining whether there are momentum effects
ANALYSIS - H. Comparison of fourth downs and third downs
ANALYSIS - I. Miscellaneous analysis:
- -- Raw numbers on success and average yards gained on third downs outside the 17
- -- Success rates on third downs for various goals to go
- -- Reestimating the V's with games with large point spreads excluded

DATA - A.

7/13/2001

WHAT TO RECORD

General:

-- Do things in EXCEL (not LOTUS).

-- Do everything in all capital letters.

-- Use Travis's abbreviations for each team.

-- By a "situation," I mean: a first and 10 (or first and goal at or inside the 10); or a kickoff from the 30 (at the beginning of a half, or following a field goal or touchdown); or a kickoff from the 20 following a safety.

-- I want to know about each situation in the first quarter (plus, as described below, a little information about the first situation in the second quarter).

-- Do each season as a separate file, each week as a separate sheet in the file.

The spreadsheet will have 1 row for each situation, and 9 columns:

Column 1: <the number of the game (1, 2, ...)>

Column 2: <abbreviation of the team with the ball>

Column 3: <an abbreviation for the situation: FD for first and 10 (or first and goal at or inside the 10); KO for kickoff from the 30; FK (for "free kick") for kickoff from the 20 following a safety>

Column 4: <abbreviation of the team whose half of the field the ball is on (for the 50-yard line, either abbreviation is fine)>

Column 5: <the yard line the ball is on>

Column 6: <the number of points scored. A negative number means that the team that did not start with the ball scored. Record all touchdowns as 7's, even if the team scored 0 or 2 on the extra point.>

Column 7: <what happened. If there is a score before the next situation, write "O" or "D" depending on whether it was the team with the ball in the situation that scored or the other team, followed by "TD", "FG", or "S" depending on whether the score was a touchdown, field goal, or safety. If the next situation is a first down, write "OFD" or "DFD" depending on which team has the first down.>

Column 8: <if the initial situation (that is, the situation described at the beginning of the line) is a first down, record whether the play on first down was a run ("R") or pass ("P"). RUNS BY THE QUARTERBACK COUNT AS PASSES! For cases that are complicated or unclear, write "R", "P", or "O" (for "other") followed by a number. Use Travis's numbering system where possible. When you encounter something you're not sure about, ask both Travis and me for advice. And if you end up creating a new number, let both of us know.>

For each game, the last row should describe the first situation in the <u>second</u>

quarter. Only fill in Columns 1-5. (We need this so that we know the outcome of the last situation in the first quarter.)

Also, write up separately the basic information about each game in an EXCEL file, with one row per game, and 6 columns: year (always "1998" for now), week number, game number, away team abbreviation, home team abbreviation, date in mm/dd/yyyy form.

Some clarifying notes:

-- On a kickoff, the team with the ball in the situation is the kicking team. Thus, for example, if the receiving team ends up with a first and 10, record "DFD" for what happened.

-- Something like a first and 15 or a kickoff from the 45 (because of a penalty) does not count as a situation.

-- Some of the information I am asking you to record is redundant. This is deliberate.

-- I don't know if it is possible to have a kickoff from the 20 after a touchdown or field goal because of penalties. If this happens, it does not count as a situation.

More on runs versus passes:

-- The goal is to determine the team's intent. Classify passes, sacks, and runs by the quarterback as passes. A play that is nullified by a penalty should still be classified depending on what it was (what play is called on the subsequent play would be irrelevant).

-- Some of Travis's numbering is unnecessary. You needn't bother with K1 and K2; you can code his P1, P3, P4, P5, and P8 as just P; and you can code his R1, R2, R4, R5, and R7 as just R. But do code his K5, P6, P7, R3, R6, R8, R9, O1, O2, O3, and O4 the same way he did. (K5 and R6 are just for future reference. And the distinction between O3 and O4 is not important, but having started it, I'd like to keep it.) WHEN IN DOUBT, ASK!

-- FINALLY, DON'T FORGET THAT RUNS BY THE QUARTERBACK COUNT AS PASSES!

DATA - A.

9/27/01
[corrected 11/5/01]

DATA MANIPULATION

We want to create a single monster spreadsheet that contains the same information as the separate spreadsheets, but in a more usable form. Specifically:

A. We want a spreadsheet with many rows, one for each situation in our sample. (Do not include rows for the second-quarter situations.) There will be 9 columns:

Column 1: The season.
Column 2: The week number.
Column 3: The game number. (Use the same numbering as on the undergraduates' spreadsheets.)
Column 4: Abbreviation of the team with the ball. (Identical to the undergraduates' spreadsheets, Column 2.)
Column 5: A number identifying the situation: "1" = first and 10 on one's own 1, "2" = first and 10 on one's own 2, ..., "99" = first and goal on the opponent's 1, "100" = kickoff from the 30, "101" = kickoff from the 20.
Column 6: Net number of points scored before the next situation. CHANGE ALL TOUCHDOWNS TO 6.984. (Identical to the undergraduates' spreadsheets, Column 7, except that the 7's need to be changed to 6.984's.)
Column 7: Dummy that equals 1 if the team with the ball in this situation also has the ball in the next situation in the game, and that equals -1 if the other team has the ball.
Column 8: A number identifying the next situation (1 through 101).
Column 9: "P" for pass, "R" for run, etc. (Identical to the undergraduates' spreadsheets, Column 9.)

DATA - B.

[c. 8/04]

I want to create a single spreadsheet with all the fourth-down data. The idea is to take the information in the spreadsheets that the four of you have been creating, combine it, and get it into a form that's easier to analyze.

Each fourth-down play will be a separate row.

There will be sixteen columns:

Column A: The season (1998, 1999, or 2000).
Column B: The week number.
Column C: The game number. (Identical to the entry in the spreadsheets you're starting from, Column A.)
Column D: The quarter. (Identical to the entry in the spreadsheets you're starting from, Column B.)
Column E-F: These two columns will show the time remaining in the quarter.
-- Column E should show just the time. That is, it should have the entry in the spreadsheets you're starting from, Column C, without any "Q", "X", or "A".
-- Column F: If there is a "Q", "X", or "A", that should go in Column F; otherwise, Column F should be blank.
Columns G-J: Identical to the entries in the spreadsheets you're starting from, Columns D-G. (The only issue here is that I want to be consistent about the team abbreviations. That is, a given team should always be abbreviated the same way. If the abbreviations are not currently consistent, change them to make them consistent.)
Column K: A number identifying where the ball is: "1" = own 1, "2" = own 2, ..., "99" = opponent's 1. This will use the information in the spreadsheets you're starting from, Columns D, H, and I. For example, if SF has the ball on the NO 27, the entry in this column should be 73. If SF has the ball on the SF 33, the entry in this column should be 33.
Column L: The number of yards the team has to go for a first down (or touchdown). This should be identical to the entries in the spreadsheets you're starting from, Column J, except that you should change all "1"'s to "0.75"'s (on the grounds that they round any number of yards to go between 0 and 1.5 to 1).
Column M: What the team tried to do: "F" for field goal, "U" for punt, "G" for going for a first down or touchdown, "O" for can't tell, with numbers and *'s added to reflect special cases and consecutive 4th downs. (Identical to the spreadsheets you're starting from, Column K.)
Column N: Net number of points scored before the next situation. (Identical to the spreadsheets you're starting from, Column L, **except that 7's should be changed to 6.984's, and -7's should be changed to -6.984's**.)
Column O: Dummy variable that equals 1 if the team with the ball on the fourth down also has the ball in the next situation in the game, and that equals -1 if the other team has the ball. This will use the information in Columns D and M of the spreadsheets you're starting from.
Column P: A number identifying the next situation (1 through 101). Situations are defined as follows: "1" = first and 10 on one's own 1,

"2" = first and 10 on one's own 2, ..., "99" = first and goal on the opponent's 1, "100" = kickoff from the 30, "101" = kickoff from the 20. This will use the information in Columns M-P of the spreadsheets you're starting from.

I've also pasted below my earlier instructions to the 4 of you about the fourth-down spreadsheets. I changed the column numbers to letters to be consistent with EXCEL.

Thanks!

D.R.

****
OLD INSTRUCTIONS:

General:
-- Do things in EXCEL (not LOTUS, etc.).
-- Do everything in all capital letters.
-- As I use it, "situation" is a technical term. It means: A first and 10 (or first and goal at or inside the 10); or a kickoff from the 30 (at the beginning of a half, or following a field goal or touchdown); or a kickoff from the 20 following a safety.
-- Note: A kickoff at the beginning of a half or following a field goal or touchdown that is not from the 30 (because of a penalty) is NOT a situation. Example: Team goes for it on fourth and goal and scores, but because of a penalty kicks off from the 45; the kickoff goes for a touchback. In this case, the next situation after the fourth down play is the first and 10 on the 20.

Each fourth-down play gets a separate row. There will be 16 columns.

Columns A-J describe the setting:

Column A: <the number of the game (1, 2, ...)>
Column B: <the quarter>
Column C: <time remaining in the quarter>
-- For plays not in the last two minutes of a half: Use the time the possession started.
-- Note #1: If the printout shows both the time the possession started and the time of the first play from scrimmage, use the time the possession started.
-- Note #2: If the possession started in the previous quarter, enter "15:00Q".
-- For plays in the last two minutes of a half: use the exact time, and put an "X" after the time.
-- Note: If the play is in the last two minutes but the exact time is not available, make a reasonable guess. For example, you could average the times of the play before and the play after. In such cases, add an "A" (for "approximate") after

the time.
Column D: <abbreviation of the team with the ball>
Column E: <the number of points the team with the ball has>
Column F: <abbreviation of the opponent>
Column G: <the number of points the opponent has>
Column H: <abbreviation of the team whose half of the field the ball is on>
-- Note: If the ball is on the 50, you can use the abbreviation of either team.
Column I: <the yard line the ball is on>
Column J: <the number of yards the team has to go for a first down (or touchdown)>

Columns K-L describe the fourth-down play:

Column K: <what the team tried to do: "F" for field goal, "U" for punt, "G" for going for a first down or touchdown. See below for an elaboration and a discussion of complications.>
Column L: <the number of points scored prior to the next situation>
-- Note #1: As always, by "situation," I mean: a first and 10, a first and goal at or inside the 10, a kickoff from the 30, or a kickoff from the 20 following a safety.
-- Note #2: A negative number means that the team that did not start with the ball scored.
-- Note #3: Record all touchdowns as 7's.

Columns M-P describe the next situation:

Column M: <abbreviation of the team with the ball in the next situation>
Column N: <abbreviation for the situation: 1D for first and 10 (or first and goal at or inside the 10), KO for kickoff from the 30, FK for kickoff from the 20 following a safety>
Column O: <abbreviation of the team whose half of the field the ball is on>
-- Note: If the ball is on the 50, you can use the abbreviation of either team.
Column P: <the yard line the ball is on>

General note on Columns M-P: If the fourth-down play if the last play of a half, Columns M-P should describe what the next situation would have been if play had continued. Example: A team tries a 31-yard field goal on a fourth down on the last play of the half. The attempt is no good. Columns M-P should describe a first and 10 for the other team on their own 31.

Complications and elaborations:

1. If there's a penalty on a fourth down that results in another fourth-down play, both plays get a row in the spreadsheet. Put a "*" after the "U," "F," or "G" for the play with the penalty, so we have a record of when this occurred.

2. In Column K, I want to know the team's intent on punt vs. field goal vs. going for it, not what actually happened. So if the team planned to do

something but didn't (because of a penalty, a blown play, or whatever), assign it a "U," "F," or "G" according to what was planned.

In some cases, the only way to tell the team's intent is to look at who a penalty was called on: if a penalty is called on the quarterback, assume the team was going for it; if on the punter, assume the team was punting; if the field-goal kicker, assume the team was trying a field goal.

Use numbers to show special or ambiguous cases, including fake punts and fake field goals and cases where the team planned to kick but ended up with a first down (because of a penalty, for example). And use "O" followed by a number if you can't tell the team's intent. Here are the numbers we have so far:

O1 PENALTY ON OFFENSE CAUSES NEW 4TH DOWN, G VS. K UNKNOWN
O2 PENALTY ON DEFENSE CAUSES NEW 1ST DOWN, G VS. K UNKNOWN
O3 PENALTY ON DEFENSE CAUSES NEW 4TH DOWN, G VS. K UNKNOWN
F1 PENALTY ON FIELD GOAL ATTEMPT CAUSES FIRST DOWN
F2 FAKE FIELD GOAL ATTEMPT
F3 FAKE FIELD GOAL ATTEMPT --> PUNT
F5 PROBABLY A FAKE, BUT DESCRIPTION UNCLEAR
U1 PENALTY ON PUNT CAUSES FIRST DOWN
U2 FAKE PUNT
U3 PENALTY ON PUNTER (NEW 4TH DOWN, FIRST WAS A PUNT)
U4 OFFSETTING PENALTIES--NEW 4TH DOWN, SAME YARD LINE
U5 FUMBLED SNAP ON PUNT; RUN (CAN'T TELL WHETHER WAS INTENDED AS A FAKE)
U6 PUNTER SENT IN; PENALTY BEFORE PLAY
U7 PUNTER SENT IN; PENALTY ON DEFENSE BEFORE THE PLAY; 1ST DOWN

3. A letter can have both a number and a "*."

DATA - C.

1/28/02

We want a spreadsheet with many rows, one for each third down in our sample. There will be 11 columns.

Column 1: The season.
Column 2: The week number.
Column 3: The game number. (Identical to undergraduates' spreadsheets, Column 1.)
Column 4: Abbreviation of the team with the ball. (Identical to undergraduates' spreadsheets, Column 2.)
Column 5: A number identifying where the ball is: "1" = own 1, "2" = own 2, ..., "99" = opponent's 1.
Column 6: The number of yards the team has to go for a first down (or touchdown). This should be identical to the undergraduates' spreadsheets, Column 5, except that you should change all "1"'s to "0.75"'s (on the grounds that they round any number of yards to go between 0 and 1.5 to 1).
Column 7: What the team tried to do: "R" for run, "P" for pass, "O" for can't tell whether run or pass, and "OO" for we wouldn't even know if the team was planning to kick (which of course they wouldn't actually do on third down), all with numbers added for special cases. (Identical to undergraduates' spreadsheets, Column 6.)
Column 8: Net number of points scored before the next fourth down or the next situation, whichever comes first. (Identical to undergraduates' spreadsheets, Column 7, <u>except that 7's should be changed to 6.984's, and -7's should be changed to -6.984's</u>.)

Columns 9-10 describe what the next situation would have been if the third-down play had taken place on fourth down. Thus, for example, if SF has third and 2 on its own 40 and throws an incomplete pass, the undergraduate's spreadsheet will have "SF," "4D," "SF," "40" in Columns 8-11. But we would want "-1" (other team has the ball) and "60" (location is <u>opponent's</u> 40) in your Columns 9-10. Specifically:

Column 9: Dummy that equals 1 if the team with the ball on the third down would also have had the ball in what the next situation would have been if the third-down play had taken place on fourth down, -1 if the other team would have had the ball.
Column 10: A number identifying what the next situation would have been if the third-down play had taken place on fourth down (1 through 101). Situations are defined as before: "1" = first and 10 on one's own 1, "2" = first and 10 on one's own 2, ..., "99" = first and goal on the opponent's 1, "100" = kickoff from the 30, "101" = kickoff from the 20.

Finally, Column 11 is a dummy to allow us to distinguish turnovers from plays where the team would have turned the ball over on downs if the play had been on fourth down. Specifically:

Column 11: A dummy equal to 1 if the entry in Column 9 of the undergraduate's

spreadsheet is "4D", 0 otherwise.

My earlier instructions about third downs:

Third downs are like fourth downs: we need to know what the circumstances were, what the team tried to do, and what happened.

Specifically:

Let's organize things with one EXCEL file for each season (or part season), with each week as a separate sheet in the file.

Each third-down play in the first quarter should get a separate row. There will be 11 columns.

Columns 1-7 describe the third-down play:

Column 1: <the number of the game (1, 2, ...)>
Column 2: <abbreviation of the team with the ball>
Column 3: <abbreviation of the team whose half of the field the ball is on>
Column 4: <the yard line the ball is on>
Column 5: <the number of yards the team has to go for a first down (or touchdown)>
Column 6: <what the team tried to do: "R" for run, "P" for pass. With one exception, use the same rules and numbering system as for first downs (see below).>
Column 7: <the number of points scored prior to the next fourth down or the next situation, whichever comes first. (By "situation," I mean the same thing as before: a first and 10, a first and goal at or inside the 10, a kickoff from the 30, or a kickoff from the 20 following a safety.) A negative number means that the team that did not start with the ball scored. Continue to record all touchdowns as 7's.>

Columns 8-11 describe the next situation or fourth down (whichever comes first):

Column 8: <abbreviation of the team with the ball in the next situation or fourth down>
Column 9: <abbreviation for the situation/fourth down. Let's use 4D for a fourth down, 1D for first and 10 (or first and goal at or inside the 10). As always, KO is a kickoff from the 30, FK is a kickoff from the 20 following a safety>
Column 10: <abbreviation of the team whose half of the field the ball is on>
Column 11: <the yard line the ball is on>

Complications and elaborations:

1. On runs versus passes, for the most part use the same rules and numbering system as for first downs. The exception is that I want to know about cases where it would have been impossible to tell if the team was planning to kick. I realize that teams never actually kick on third down, but we will probably be excluding plays like this from our analysis of fourth downs. To avoid biasing the results, we will probably also want to exclude them from our analysis of third downs. The one case I can think of where we would not be able to tell if the team was planning to kick is a penalty called before the snap on someone other than the quarterback. Use "O" for cases where you can tell that the team was not planning to kick but can't tell whether it was planning to run or pass (for example, a fumble by the QB on the snap), and "OO" for cases where you wouldn't have been able to tell if it was planning to kick, with appropriate numbers afterward. For 2000, the only OO's were "OO1 -- Defensive penalty, no play information" and "OO2 -- Offensive penalty, no play information."

Otherwise, use the same numbering system as on first downs. And, of course, runs by the quarterback count as passes.

2. If there's a penalty on a third down that results in another third-down play, both plays get a row in the spreadsheet. Put a "*" after the "R," "P," "O," or "OO" for the play with the penalty, so we have a record of when this occurred.

3. I believe that a third-down play is always followed by a fourth-down play, a "situation," or another third-down play. Let me know if a third-down play is ever immediately followed by something else.

4. Likewise, let me know if a team ever tries to punt or kick a field goal on third down.

DATA - D.

[7/9/01]

> c. While it's still fresh, could you write me a few sentences
>about these data (what they are [which casino, etc.] and where you got
>them [web address, date accessed if you know it, etc.])?

The data are from the following web address:

http://sports.dbc.com/cgi-bin/htx.exe/pastscores/archives/2000/files/0901lin
.txt?source=pastscores/psvi/2000/september

accessed on June 24th. First I selected the proper month in the dropdown box, then I clicked on the date on the calendar at the top so that I could get each game on the day it was played (although two days I had to get the day before the game because the game day was missing-- Dec 18th and Dec 24th. When I checked the "day before" data on several other days, it was the same as the day of data, it simply did not have the final outcome listed.). I used the "Current" column of the Stardust casino data.

DATA - D.

3/26/01

I would like to know what fractions of 1-point and of 2-point extra-point attempts are successful. There are two complications. First, I want any broken plays on 1-point attempts where the team unintentionally tried for 2 points (because of a bad snap, for example) classified with the 1-point attempts. Second, if there are any fake 1-point attempts, I want them classified separately from both the regular 1-point attempts and the regular 2-point attempts.

I would like this information for all four quarters. I think what would be best would be a spreadsheet with one row per week and 11 columns:

Column 1: Week number
Column 2: Number of 1-point tries
Column 3: Number of 1-point tries that produced 1 point
Column 4: Number of 1-point tries that produced 0 points
Column 5: Number of 1-point tries that produced 2 points
Column 6: Number of 2-point tries
Column 7: Number of 2-point tries that produced 2 points
Column 8: Number of 2-point tries that produced 0 points
Column 9: Number of fake 1-point tries
Column 10: Number of fake 1-point tries that produced 2 points
Column 11: Number of fake 1-point tries that produced 0 points

(Columns 3+4+5 should equal Column 2; 7+8 should equal 6; and 10+11 should equal 9.)

To do this, you need to go through all four quarters of each game and find each touchdown, keep a tally of what happened on the extra points, and enter the tally for each week.

ANALYSIS - A.

11/7/01

ESTIMATING THE V'S

A. We want to estimate the unconstrained (that is, unsplined) equation (4) in the paper by the IV procedure described there.

B. We want to estimate the constrained (splined) version. Let's have knot points at the 9, 17, and 33 yard lines for both teams, and the 50 (7 knot points, so the field is divided into 8 intervals).

If what follows seems mysterious or mistaken (or both), let me know, and I'll try to explain what I was thinking.

Let A_y, B_y, and C_y be vectors that pick up the constant, slope, and quadratic terms for each interval between knots. Each vector will have 101 elements (for the 99 yard lines, plus kickoffs and free kicks). The y subscripts are for the yard line at the beginning of the interval.

Concretely:

A_1 is a vector of 8 1's (for the 1 through 8 yard lines) followed by 93 0's.
A_9 is a vector of 8 0's, then 8 1's, then 85 0's.
A_{17} is a vector of 16 0's, then 16 1's, then 69 0's.
...
A_{91} is a vector of 90 0's, then 9 1's, then 2 0's.
A_{100} and A_{101} each have only one 1: A_{100} has it in the 100th entry, and A_{101} has it in the 101st.

B_1 is a vector consisting of 0, 1, 2, 3, 4, 5, 6, 7 (for the first 8 yard lines), followed by 93 0's.
B_9 is a vector of 8 0's, then 0, 1, 2, 3, 4, 5, 6, 7, then 85 0's.
B_{17} is a vector of 16 0's, then 0, 1, 2, ..., 15, then 69 0's.
...
B_{91} is a vector of 90 0's, then 0, 1, 2, ..., 8, then 2 0's.
There is no B_{100} or B_{101}.

C_1 is a vector consisting of 0, 1, 4, 9, 16, 25, 36, 49 (for the first 8 yard lines), followed by 93 0's.
C_9 is a vector of 8 0's, then 0, 1, 4, ..., 49, then 85 0's.
C_{17} is of vector of 16 0's, then 0, 1, 4, ..., 225, then 69 0's.
...
C_{91} is a vector of 90 0's, then 0, 1, 4, ..., 64, then 2 0's.
There is no C_{100} or C_{101}.

I'll leave it up to you whether there's some clever matrix or programming way to create the A's, B's, and C's, or whether to just do it by brute force.

Use these to construct new variables. Let X_{gt} be a vector of the 101 X^i_{gt}'s.

Define $X_{gt}^{A1} = A_1' X_{gt}$ (so X_{gt}^{A1} is a scalar, not a vector), and so on for all the A's, B's, and C's. This is a total of 26 variables.

Now do the estimation. In the unconstrained version, we've estimated

$$P_{gt} = \sum_i V_i X_{gt}^i + e_{gt}, \tag{4}$$

instrumenting with the 101 D^i's. The constrained version does two things:

First, it replaces the 101 X^i's with the 26 new variables you've just constructed, still instrumenting with the 101 D^i's. If I've done things right, this forces the V's over each interval to follow a quadratic.

Second, it places restrictions on the coefficients on the new variables. Let L_y denote the coefficient on X^{Ay}, S_y the coefficient on X^{By}, and Q_y the coefficient on X^{Cy}. Then there are 2 sets of restrictions:

-- Level restrictions: $L_1 + 8S_1 + 64Q_1 = L_9$, $L_9 + 8S_5 + 64Q_5 = L_{17}$, $L_{17} + 16S_{17} + 256Q_{17} = L_{33}$, ..., $L_{83} + 8S_{83} + 64Q_{83} = L_{91}$ (7 restrictions).

-- Slope restrictions: $S_1 + 16Q_1 = S_9$, $S_9 + 16Q_9 = S_{17}$, $S_{17} + 32Q_{17} = S_{33}$, ..., $S_{83} + 16Q_{83} = S_{91}$ (7 more restrictions).

C. I'd like to see output files for both the unconstrained and constrained estimation, and plots of V_1 through V_{99} for both methods. (With the spline, the V's are functions of the L_y's, S_y's, and Q_y's -- see Section IV below.)

IV. STANDARD ERRORS FOR THE V'S IN THE SPLINED CASE

Let $\hat{a}$ denote the estimated vector of coefficients from the splined estimation in III.B, and let $\hat{\Omega}$ denote the estimated variance-covariance matrix of a. (Thus a is 26 x 1 and Ω is 26 x 26. The standard error of the ith element of a is the square root of the ith diagonal element of $\hat{\Omega}$.)

Now consider a given $\hat{V}_i$. $\hat{V}_i$ is a linear function of the elements of $\hat{a}$. We can write this as $\hat{V}_i = c_i'\hat{a}$, where c_i' is a vector telling us how to find $\hat{V}_i$ from $\hat{a}$. The standard error of $\hat{V}_i$ is the square root of $c_i'\hat{\Omega}c_i$.

I don't know what the easiest way to do this is. Any package should have an easy way to get the point estimate and standard error for a linear combination of coefficients. Thus you could just have 101 instructions asking for such estimates and standard errors (one for each V_i). Or you could stack the 101 c_i''s to produce a 101 x 26 matrix C' and then find $C'\hat{\Omega}C$. This would be the full 101 x 101 variance-covariance matrix of the $\hat{V}_i$'s; their standard errors would be the square roots of the diagonal elements. I have a slight preference for the $C'\hat{\Omega}C$ approach, but I don't feel strongly.

We will also use $\hat{a}$ and $\hat{\Omega}$ later to get correct standard errors for other things we are doing.

ANALYSIS - B.

12/6/04

III. THE VALUE OF KICKING

A. OVERVIEW

Because of corrections to the data, we need to redo the estimation of the value of kicking. Much of what follows is lifted from my old notes.

Where we are heading is estimating equation (5) in the paper:

$$\tilde{P}_{gt} + \tilde{B}_{gt}\sum_i \tilde{D}^i_{gt}V_i = \sum_i K_i A^i_{gt} + u_{gt}. \qquad (5)$$

The paper's explanation of the notation is:

> Let K_i denote the value of a kick from yard line i, and let t index kicks within a game. In addition, let A^i_{gt} be a dummy that equals 1 if kick g,t is from yard line i, $\tilde{D}^i_{gt}$ a dummy that equals 1 if the next situation after kick g,t is a situation of type i, $\tilde{P}_{gt}$ the net number of points scored by the kicking team between kick g,t and the subsequent situation (including any points scored on the kick itself), and $\tilde{B}_{gt}$ a dummy that equals 1 if the kicking team has the ball in the next situation and -1 if the other team has the ball.

The V_i's are the values of the different situations (from what we've already done).

B. THE SAMPLE AND BASIC INFORMATION

Use only first-quarter observations!

Exclude "G"'s, "O1"'s, "O2"'s, and "U4"'s. This should leave only various kinds of "F"'s and "U"'s in the sample. As before, include fake kicks ("F2"'s, "F4"'s, "F5's", and "U2"'s). Likewise, leave in the "F3"'s.

Let me know how many of each of "F2"'s, "F4"'s, "F5"'s, "U2"'s, "U5's", and "F3"'s there are.

I would like to know how many observations there are in the sample where the kicking team had 5 or fewer yards to go, and on how many of these observations the kicking team moved 5 or fewer yards and made a first down or touchdown.

I would like to know the total number of observations in the sample.

C. THE ESTIMATION

1. BASICS

Estimate equation (5).

Throughout, use the V's from our preferred spline estimation (knot points at the 9's, 17's, 33's, and 50).

As in the current version, the K_i's should be splined, with the same knot points as the V_i's, but not imposing the <u>slope</u> restriction at the opponent's 33 (do impose the levels restriction, however).

2. HOW TO REPORT THE RESULTS

2 ways:

-- Simply report the estimated K_i's.

-- Report the <u>differences</u> between the estimated K_i's and the values of the other team having a first down on the spot. This is $K_i + V_{i'}$, where i' is the "opposite" yard line from i (e.g., if i = own 20, then i' = opponent's 20).

D. STANDARD ERRORS

As before, we'll want standard errors for both the K_i's and the $K_i + V_i$'s. What follows is cribbed from my old notes.

There are two complications: the splining, and the fact that the V_i's are estimated.

Let's again use $\hat{a}$ to denote the estimated vector of coefficients from the splined estimation of the V's, and $\hat{\Omega}$ the estimated variance-covariance matrix of a. Let N_a denote the number of elements of a. So a is N_a x 1 and Ω is N_a x N_a. By my count, N_a = 26.

Let's use $\hat{b}$ to denote the estimated vector of coefficients from the estimation of the K's, and $\hat{\lambda}$ denote the estimated variance-covariance matrix of b neglecting the fact that the V's are estimated. Let N_b denote the number of elements of b. By my count, N_b = 24.

I propose to first find the variance-covariance matrix of the $(N_a + N_b)$ x 1 vector [a,b]. I think it is:

$$\begin{bmatrix} \hat{\Omega} & [[db/da]\hat{\Omega}]' \\ [db/da]\hat{\Omega} & \hat{\lambda} + [db/da]\hat{\Omega}[db/da]' \end{bmatrix} \equiv \hat{\Gamma}$$

Here db/da is an N_b x N_a matrix whose ij element is the derivative of b_i with respect to a_j. Ω is N_a x N_a, $[db/da]\hat{\Omega}$ is N_b x N_a, $[[db/da]\hat{\Omega}]'$ is N_a x N_b, and $\hat{\lambda} + [db/da]\hat{\Omega}[db/da]'$ is N_b x N_b.

Once we have $\hat{\Gamma}$, we can find the things that we need. Suppose there is something, ϕ, that is a linear function of [a,b]: $\phi = d'[a,b]$, where d is a $(N_a + N_b)$ x 1 vector. Then the standard error of ϕ is the square root of $d'\hat{\Gamma}d$.

Each K_i and each $K_i - V_i$, takes the form d'[a,b]. Each K_i does not depend directly on the a's, and so the d vector for each K_i will have zeroes for the first N_a elements. As with the V_i's, it may be easier to stack the 99 d_i's for the 99 K_i's to produce a 99 x $(N_a + N_b)$ matrix D' and then find $D'\hat{\Gamma}D$. Each $K_i - V_i$ depends on both a and b (K_i depends on b and V_i depends on a), but otherwise the logic is the same.

The calculation of db/da will have to be done numerically. Here's one way to do it:

i. Do our usual stuff, and find $\hat{b}$.

ii. Add 0.001 to the first element of $\hat{a}$; then redo the estimation of b. Call the new estimate $\hat{b}^*$.

iii. Then the first column of $\partial\hat{b}/\partial\hat{a}$ is $(\hat{b}^* - \hat{b})/0.001$.

iv. Restore the first element of $\hat{a}$ to where it started.

v. Repeat steps (ii)-(iv) for the second element of a, the third element, and so on.

vi. Also compute the derivatives subtracting rather than adding 0.001 from each element of $\hat{a}$. Each derivative should be close to the corresponding one the first time around. If not, this will be a sign that our numerical derivatives may differ in an important way from the true derivatives.

ANALYSIS - C.

1/28/02

II. THE VALUE OF GOING FOR IT

We want to estimate the value of going for it on fourth down. The main difference compared with your analysis of the value of kicks is that the value of going for it depends not only on the team's position on the field, but on the number of yards to go for a first down or touchdown.

One of the packages that Ryan used for this stuff had a bug in it that gave us insane results. So you might want to talk to Ryan about what package to avoid.

A. SAMPLES AND BASIC INFORMATION

1. Samples

I would like to consider two samples. One excludes only the "OO"'s (including "OO"'s with numbers and/or "*"'s after them). The other also excludes observations where the team has 11 or more yards to go.

2. Basic information

a. I would like to know the number of observations in each sample.

b. Considering only observations between one's own 10 and the opponent's 20 (inclusive), I'd like to know, for each number of yards to go (1 yard to go, 2 yards to go, ..., up through 15): (1) the number of observations with that number of yards to go; (2) the number of those times that the team would have made it (that is, the number of those times when either the entry in Column 9 of your spreadsheet is a 1 or the entry in Column 8 is strictly positive); (3) the average number of yards the ball moved forward (from the perspective of the team with the ball). (For (3), whenever the entry in Column 8 of your spreadsheet is 0, you can use the information in Columns 9-10 to figure out where the ball ended up. When the entry in Column 8 is not 0, compute the number of yards the ball moved as the number of yards to the relevant goal line -- the opponent's goal line if the number of points scored is positive, the team's own goal line if the number of points scored is negative.)

c. Considering only observations where the team had goal to go, I'd like to know, for each number of yards to go (3rd and goal on the 1, 3rd and goal on the 2, ..., up through 10): (1) the number of observations with that number of yards to go; (2) the number of those times that the team made it (that is, the number of those times when the entry in Column 8 of your spreadsheet is a 6.984); (3) the number of times where the team didn't make it but would have retained the ball (that is, Column 8 = 0 and Column 9 = 1); (4) the average number of yards the ball moved forward. (Computed in the same way as in (b).)

TO BE CONTINUED ...

ANALYSIS - C.

2/20/02

Here are some more third-down instructions; they pick up where the previous batch leaves off. They will allow us to finally address the questions that motivate the paper: When should teams go for it on fourth down? How do their actual choices compare with what they should do? How large are the gains from making better choices? Even more so than usual, I've probably erred on the side of spelling things out in gory detail.

I suggest reading pp. 11-15 of the paper ("Going for It" through the end of Section III) before tackling these notes.

B. THE DEPENDENT VARIABLE

The dependent variable is what the realized value of going for it would have been if the play had been on fourth down. That is (borrowing notation from the fourth-down analysis), it is $\tilde{P}_{gt} + \tilde{B}_{gt}\sum_i \tilde{D}^i_{gt} V_i$, where:

-- $\tilde{P}_{gt}$ is the net number of points scored by the team with the ball before the next fourth down or the next situation (Column 8 from your spreadsheet);

-- $\tilde{B}_{gt}$ is a dummy that equals 1 if the team would have retained the ball and -1 if the other team would have gotten the ball (Column 9 from your spreadsheet);

-- $\tilde{D}^i_{gt}$ (for i = 1 to 101) is a dummy that equals 1 if the next situation would have been a situation of type i (so $\tilde{D}^i = 1$ if and only if the entry in Column 10 of your spreadsheet equals i).

-- The V_i's are the estimated values of the different situations (from what we've already done). Continue to use the V_i's from our preferred spline estimation (knot points at the 9's, 17's, 33's, and 50).

Let's use Y_{gt} to denote the dependent variable.

C. THE ESTIMATION

1. The basic idea

Look at equation (6) in the paper, and the discussion around it. I want to make two changes relative to that material. First, I want to make the estimates quadratic rather than cubic in the number of yards to go. Thus the modified equation (6) is:

$$G_{py} = a_1 + a_2 y + a_3 p + a_4 y^2 + a_5 yp + a_6 p^2 + a_7 y^2 p + a_8 yp^2 + a_9 y^2 p^2. \qquad (6')$$

Second, let's divide things at the usual yard lines (9's, 17's, 33's, and 50).

(6') has 9 right-hand-side variables (counting the constant), and the coefficients on them (the a_k's) are allowed to be different in each "slice" (with some constraints). So we need to create a bunch of variables, Z^{jk}, where j indexes slices and k runs from 1 to 9. Z^{jk} captures the kth right-

hand-side variable of (6') for slice j. By my count, there are 8 x 9 = 72 Z's. We will then run

$$Y_{gt} = \sum_j \sum_k a_{jk} Z_{gt}^{jk} + u_{gt}.$$

We will impose constraints on the a's so that the level and derivative of the value of going for it are continuous along the "seams."

2. The Z's

Let A be a vector of 99 elements: 0, 1, 2, 3, 4, 5, 6, 7; then 0, 1, 2, 3, 4, 5, 6, 7 again; then 0, 1, 2, ..., 15; then 0, 1, 2, ..., 16; then 0, 1, 2, ..., 16 again; then 0, 1, 2, ..., 15; then 0, 1, 2, 3, 4, 5, 6, 7; then 0, 1, 2, 3, 4, 5, 6, 7, 8. The ith element of A shows the number of yards that the ith yard line is from the beginning of its slice. For example, the opponent's 25 is 8 yards from the beginning of its slice (since there is a slice starting at the opponent's 33 and going to its 17), so the 75th element of A is 8.

Let B_{gt}^1 through B_{gt}^8 be dummies for what slice third down g,t is in. That is, $B_{gt}^1 = 1$ iff third down g,t is between the team's 1 and 8 (inclusive); $B_{gt}^2 = 1$ iff third down g,t is between the team's 9 and 16 (inclusive); ... ; $B_{gt}^8 = 1$ iff third down g,t is between the opponent's 9 and 1 (inclusive).

Let y_{gt} be the number of yards the team has to go on third down g,t.

Let p_{gt} be the ith element of A, where i is the entry in Column 5 of your spreadsheet for that observation. (Thus p_{gt} is a scalar; it's the number of yards the team is from the beginning of the slice it is in. This corresponds to p in equation (6) in the paper.)

Now, for j = 1 , ..., 8, let:

$$Z_{gt}^{j1} = B_{gt}^j$$

$$Z_{gt}^{j2} = B_{gt}^j y_{gt}$$

$$Z_{gt}^{j3} = B_{gt}^j p_{gt}$$

$$Z_{gt}^{j4} = B_{gt}^j (y_{gt})^2$$

$$Z_{gt}^{j5} = B_{gt}^j y_{gt} p_{gt}$$

$$Z_{gt}^{j6} = B_{gt}^j (p_{gt})^2$$

$$Z_{gt}^{j7} = B_{gt}^j (y_{gt})^2 p_{gt}$$

$$Z_{gt}^{j8} = B_{gt}^j y_{gt} (p_{gt})^2$$

$$Z_{gt}^{j9} = B_{gt}^j (y_{gt})^2 (p_{gt})^2$$

(So 8 x 9 = 72 Z's.)

3. Explanation

You're probably wondering what the hell this is all about. Look at equation (6'). The idea of the above is that each j corresponds to one of our 8 slices, and the Z^{jk}'s for k=1,...,9 pick up the 9 terms in (6') for that slice. In each case, p is measured relative to the beginning of whatever slice the observation is in (that's what the A vector is doing).

I'd like you to check this by choosing one observation and checking that the Z^{jk}'s are right for that observation. Say you pick an observation where there is a third and 10 on the team's own 38. This is in slice 4, and y = 10 (10 yards to go), p = 5 (5 yards from the start of the slice). So for this observation, we should have:

-- For $j \neq 4$, all the Z^{jk}'s should equal zero for this observation.

-- For the Z^{4k}'s, we should have $Z^{41} = 1$, $Z^{42} = 10$, $Z^{43} = 5$, $Z^{44} = 100$, $Z^{45} = 50$, $Z^{46} = 25$, $Z^{47} = 500$, $Z^{48} = 250$, $Z^{49} = 2500$.

4. The unconstrained regression

This is easy (see "the basic idea," above):

$$Y_{gt} = \sum_j \sum_k a_{jk} Z^{jk}_{gt} + u_{gt}.$$

5. The constraints

Define $\bar{p}_j$ as the width of slice j. Thus $\bar{p}_1 = 8$, $\bar{p}_2 = 8$, $\bar{p}_3 = 16$, $\bar{p}_4 = \bar{p}_5 = 17$, $\bar{p}_6 = 16$, $\bar{p}_7 = 8$. ($\bar{p}_8$, which equals 8, isn't relevant.) Consider the seam where the first two slices meet, which is at the 9-yard line. In terms of slice 1, $p = \bar{p}_1$ (we are $\bar{p}_1$ yards from the start of slice 1). In terms of slice 2, p = 0 (we are at the start of slice 2). To find the level of the fitted value of the regression along the seam (that is, at the 9-yard line, for all y) from the point of view of slice 1, we use equation (6') with the a_{1k}'s and with $p = \bar{p}_1$:

$$a_{11} + a_{12}y + a_{13}\bar{p}_1 + a_{14}y^2 + a_{15}y\bar{p}_1 + a_{16}\bar{p}_1^2 + a_{17}y^2\bar{p}_1 + a_{18}y\bar{p}_1^2 + a_{19}y^2\bar{p}_1^2$$

Collecting terms gives us

$$(a_{11} + a_{13}\bar{p}_1 + a_{16}\bar{p}_1^2)$$

$$+ (a_{12} + a_{15}\bar{p}_1 + a_{18}\bar{p}_1^2)y$$

$$+ (a_{14} + a_{17}\bar{p}_1 + a_{19}\bar{p}_1^2)y^2$$

The fitted values of the regression along the seam from the point of view of slice 2 can be found from (6') using the a_{2j}'s, with p = 0. This is:

$$a_{21} + a_{22}y + a_{24}y^2$$

So we have 3 restrictions:

$$a_{11} + a_{13}\bar{p}_1 + a_{16}\bar{p}_1^2 = a_{21}$$

$$a_{12} + a_{15}\bar{p}_1 + a_{18}\bar{p}_1^2 = a_{22}$$

$$a_{14} + a_{17}\bar{p}_1 + a_{19}\bar{p}_1^2 = a_{24}$$

We can do a similar exercise for the derivative with respect to p along the seam. From the point of view of slice 1, this derivative is

$$a_{13} + a_{15}y + 2a_{16}\bar{p}_1 + a_{17}y^2 + 2a_{18}y\bar{p}_1 + 2a_{19}y^2\bar{p}_1$$

Collecting terms gives us

$$(a_{13} + 2a_{16}\bar{p}_1)$$

$$+ (a_{15} + 2a_{18}\bar{p}_1)y$$

$$+ (a_{17} + 2a_{19}\bar{p}_1)y^2$$

From the point of view of slice 2, the derivative with respect to p along the seam is simpler:

$$a_{23} + a_{25}y + a_{27}y^2$$

So we have 3 more restrictions:

$$a_{13} + 2a_{16}\bar{p}_1 = a_{23}$$

$$a_{15} + 2a_{18}\bar{p}_1 = a_{25}$$

$$a_{17} + 2a_{19}\bar{p}_1 = a_{27}$$

Repeating this analysis for the 6 other seams gives us 36 more restrictions, for a total of 42. We can now impose the restrictions and run the regression.

D. A CHECK

Let i index the team's position on the field, and y the number of yards to go. Let G_{iy} denote the estimated value of going for it in position i,y (that is, G_{iy} is the fitted value of the regression).

One would expect G_{iy} to be increasing in i and decreasing in y. I would like to know, for $0 \le y \le 20$, whether there are any (i,y) combinations such that $\partial G_{iy}/\partial i < 0$ or $\partial G_{iy}/\partial y > 0$. If there are any, let me know before proceeding.

E. REPORTING THE RESULTS

Plot G_{iy} as a function of i for y = 1, y = 5, and y = 10.

Plot the difference between the value of going for it and the value of the other team having a first and 10 on the spot as a function of i for y = 1, y = 5, and y = 10. This difference is $G_{iy} + V_{i'}$, where i' is the "opposite"

yard line from i (for example, if i is one's own 20, i' is the opponent's 20).

ANALYSIS - C.

3/21/02

I.

I want to take a different approach. Both logic and our results so far suggest that over most of the field, the _difference_ between the values of going for it and of turning the ball over on the spot should mainly depend on the number of yards to go and should be pretty much independent of where the team is on the field. So I want to focus on estimating these differences (what we've called the G{iy} + V{i'}'s), with a much more restricted specification.

So: our dependent variable will now be our old dependent variable minus the value to the team with the ball of the other team having the ball on the spot. Our old dependent variable was what the realized value of going for it would have been if the play had been on fourth down. This equaled the net number of points scored plus the value of what the next situation would have been for the team that had the ball had the play taken place on fourth down; we called this Y{gt}. Our new dependent variable is Y{gt} + V{i'}. Call it Q{gt}.

0. Let's do everything below for 3 samples -- the ones you've called 1, 2, and 3.

1. Compute the mean and standard deviation of Q{gt} over all observations in the sample.

2. Regress Q{gt} on a constant, y{gt}, and y{gt}**2, where y{gt} is the number of yards to go. Tell me the parameter estimates and their standard errors. Use the resulting variance-covariance matrix of the regression parameters to find the point estimate and standard error of G{iy} + V{i'} for y = 0, 1, 2, ..., 20.

Example: Suppose y = 10. Call the 3 regression coefficients A, B, and C, and their variance-covariance matrix E. Then G{iy} + V{i'} = A + 10*B + 100*C, and its standard error is the square root of [1 10 100]'E[1 10 100].

Since we're only estimating 3 parameters, I'm hoping for small standard errors.

3. The idea that G{iy} + V{i'} doesn't depend on i is least plausible near the goal lines. As a first pass to see if this matters, let D1{gt} be a dummy for whether observation g,t is between the 1 and the 17 (inclusive), D2{gt} be a dummy for whether observation g,t is between the 18 and the 82 (inclusive), and D3{gt} be a dummy for whether observation g,t is between the 83 and the 99 (inclusive). Now regress:

```
Q{gt} = A1*D1 + B1*D1*y{gt} + C1*D1*[y{gt}**2]
      + A2*D2 + B2*D2*y{gt} + C2*D2*[y{gt}**2]
      + A3*D3 + B3+D3*y{gt} + C3*D3*[y{gt}**2] + error
```

(This is effectively running separate regressions for the 3 separate sections of the field.) Then:

- Report the parameter estimates and their standard errors.
- Compute and plot the fitted G{iy} + V{i'}'s as a function of y for each of the 3 sections of the field (A1 + B1*y + C1*[y**2], etc.).
- Do 3 F-tests: (i) A1 = A2, B1 = B2, C1 = C2; (ii) A3 = A2, B3 = B2, C3 = C2; (iii) A1 = A2 = A3, B1 = B2 = B3, C1 = C2 = C3.

ANALYSIS - C.

4/4/02

Here's what I'm thinking.

A. Regressions

Do these for two samples: the sample that excludes observations where the team has more than 20 yards to go (I think this is "Sample 3"), and a sample that excludes observations where the team has more than 15 to go (this is a new sample).

The dependent variable will be Q{gt} (the thing you've been working with lately).

In general, the expectation of Q{gt} depends on both i (position on the field) and y (yards to go). But up through the 83, we want to make it only quadratic in y, and independent of i. Thereafter, we want to let it depend on both, in a way that is smooth. To do this, define J{gt} to be 0 if the yard line (i) corresponding to observation g,t is <= 83, then = 1 if i = 84, = 2 if i = 85, ..., = 16 if i = 99.

Then our regression is

```
Q{gt} = c{1} + c{2}*y{gt} + c{3}*[y{gt}**2] + c{4}*[J{gt}**2]
        + c{5}*y{gt}*[J{gt}**2] + c{6}*[y{gt}**2]*[J{gt}**2]
        + error
```

(no constraints needed).

I'd like to know the fitted values as a function of i and y (for y up to 20). For i up to 83, the fitted value doesn't depend on i. After i = 83, it does (via J). For example, the fitted value for i = 99, y = 10, is c{1} + c{2}*10 + c{3}*100 + c{4}*256 + c{5}*10*256 + c{6}*100*256. I'm especially interested in cases where the fitted value either is negative or is increasing in y.

Also, compute the standard error of the fitted value as a function of y (neglecting the fact that the V's are estimated) for : (i) any i <=83; (ii) i = 95.

B. Other

After this, you should parallel my earlier notes, starting with "III. WHAT SHOULD TEAMS DO?" through the end of "V. STANDARD ERRORS." The only things I saw that need to be changed are:

i. The fitted values of our new regressions give us G{iy} + V{i'}'s. So to find the G{iy}'s (which is what III-IV-V focus on), we need to subtract the V{i'}'s from the fitted values.

ii. In V.A, let's now use c^ to denote the estimated vector of coefficients from the estimation of the G{iy} + V{i'}'s., and theta^ to denote the estimated variance-covariance matrix of c neglecting the fact that the V's are estimated. (So now N{c} = 6.)

iii. Again in V.A, in the calculation of dc/dalpha: Recall that Q{gt} (our new dependent variable) equals Y{gt} (our old dependent variable) plus V{i'}. So when we change an element of alpha by 0.001, the Q{gt}'s will change both via the Y{gt}'s and via the V{i'}'s.

ANALYSIS - C.

2/28/02

This is the rest of the third-down instructions; they pick up where the previous batch leaves off.

V. STANDARD ERRORS

A. THE VARIANCE-COVARIANCE MATRIX OF THE ESTIMATED PARAMETERS

Let's again use $\hat{a}$ to denote the estimated vector of coefficients from the splined estimation of the V's, and $\hat{\Omega}$ the estimated variance-covariance matrix of a. Let N_a denote the number of elements of a. So a is N_a x 1 and Ω is N_a x N_a. By my count, $N_a = 26$.

Let's again use $\hat{b}$ to denote the estimated vector of coefficients from the estimation of the K's, and $\hat{\lambda}$ denote the estimated variance-covariance matrix of b neglecting the fact that the V's are estimated. Let N_b denote the number of elements of b. By my count, $N_b = 24$.

And let's use $\hat{c}$ to denote the estimated vector of coefficients from the estimation of the G's, and $\hat{\Theta}$ denote the estimated variance-covariance matrix of c neglecting the fact that the V's and K's are estimated. Let N_c denote the number of elements of c. By my count, $N_c = 72$.

Let's proceed in the same way as for fourth downs. So we want to first find the variance-covariance matrix of the $(N_a + N_b + N_c)$ x 1 vector [a,b,c]. To describe it, let α denote the $(N_a + N_b)$ x 1 vector [a,b]. Then the variance-covariance matrix of [a,b,c] is:

$$\begin{bmatrix} \hat{\Gamma} & [[dc/d\alpha]\hat{\Gamma}]' \\ [dc/d\alpha]\hat{\Gamma} & \hat{\Theta} + [dc/d\alpha]\hat{\Gamma}[dc/d\alpha]' \end{bmatrix} \equiv \hat{\Sigma}$$

Here $\hat{\Gamma}$ is the variance-covariance matrix of [a,b] (from our fourth-down stuff).

The calculation of $dc/d\alpha$ will have to be done numerically. Here's my proposal, which is similar to how we did the others:

i. Do our usual stuff, and find $\hat{c}$.

ii. Set j = 0.

iii. Set j = j + 1.

iv. Add 0.001 to the jth element of $\hat{\alpha}$. If this element is an element of a, recompute the V's but not the K's. If this element is an element of b, recompute the K's but not the V's.

v. Redo the estimation of the c's. Call the new estimate $\hat{c}^*$.

vi. Find the jth column of $\partial\hat{c}/\partial\hat{\alpha}$ as $(\hat{c}^* - \hat{c})/0.001$.

vii. Restore the jth element of $\hat{\alpha}$ to where it started.

viii. Go back to (iii), and repeat until you've done all the elements of α.

ix. Also compute the derivatives subtracting rather than adding 0.001 from each element of $\hat{\alpha}$. Each derivative should be close to the corresponding one the first time around. If not, this will be a sign that our numerical derivatives may differ in an important way from the true derivatives.

B. COMPUTING THE STANDARD ERRORS WE WANT

Let ω denote the vector [a,b,c]

1. The G_{iy}'s and the $G_{iy} - V_{i'}$'s

Each G_{iy} and $G_{iy} - V_{i'}$ takes the form $d'\omega$, where d is some $(N_a + N_b + N_c) \times 1$ vector. So the relevant standard error is the square root of $d'\hat{\sum}d$.

What I want for now are the standard errors of $G_{iy} - V_{i'}$ for y = 1 and y = 10, for all i. I'd like plots showing $G_{iy} - V_{i'}$ and its two-standard-error bands as functions of i for y = 1 and y = 10.

12/6/04

IV. RECOMMENDED CHOICES

Because we've changed the estimates of the value of kicking, we have to redo the analysis of recommended choices. Again, I've borrowed heavily from my old notes.

A. POINT ESTIMATES

Let's use y_i^* to denote the number of yards to go where the team should be indifferent between kicking and going for it. That is, y_i^* is defined by $G_{iy_i^*} = K_i$. You can find the y_i^*'s either using the quadratic formula (but watch out for imaginary answers) or by some type of numerical search.

Plot y_i^* as a function of i.

(Also, it doesn't make sense to consider impossible situations: (i) If we find $y_i^* < 0$, such show it as 0 in the plot. (ii) If we find y_i^* exceeds the distance to the opponent's goal line, just set it to that distance in the plot. (iii) If we find y_i^* is less than the distance to one's 10-yard line, just set it to that distance in the plot.)

12/6/04

VI. STANDARD ERRORS FOR THE y_i^*'s

This is cribbed from my old notes.

A. THE VARIANCE-COVARIANCE MATRIX OF THE ESTIMATED PARAMETERS

Let's again use $\hat{a}$ to denote the estimated vector of coefficients from the splined estimation of the V's, and $\hat{\Omega}$ the estimated variance-covariance matrix of a. Let N_a denote the number of elements of a. So a is N_a x 1 and Ω is N_a x N_a. By my count, $N_a = 26$.

Let's again use $\hat{b}$ to denote the estimated vector of coefficients from the estimation of the K's, and $\hat{\Lambda}$ denote the estimated variance-covariance matrix of b neglecting the fact that the V's are estimated. Let N_b denote the number of elements of b. By my count, $N_b = 24$.

And let's use $\hat{c}$ to denote the estimated vector of coefficients from the estimation of the G's, and $\hat{\Theta}$ denote the estimated variance-covariance matrix of c neglecting the fact that the V's and K's are estimated. Let N_c denote the number of elements of c. By my count, $N_c = 72$.

Let's proceed in the same way as for fourth downs. So we want to first find the variance-covariance matrix of the $(N_a + N_b + N_c)$ x 1 vector [a,b,c]. To describe it, let α denote the $(N_a + N_b)$ x 1 vector [a,b]. Then the variance-covariance matrix of [a,b,c] is:

$$\begin{bmatrix} \hat{\Gamma} & [[dc/d\alpha]\hat{\Gamma}]' \\ [dc/d\alpha]\hat{\Gamma} & \hat{\Theta} + [dc/d\alpha]\hat{\Gamma}[dc/d\alpha]' \end{bmatrix} \equiv \hat{\Sigma}$$

Here $\hat{\Gamma}$ is the variance-covariance matrix of [a,b] (from our fourth-down stuff).

The calculation of $dc/d\alpha$ will have to be done numerically. Here's my proposal, which is similar to how we did the others:

i. Do our usual stuff, and find $\hat{c}$.

ii. Set j = 0.

iii. Set j = j + 1.

iv. Add 0.001 to the jth element of $\hat{\alpha}$. If this element is an element of a, recompute the V's but not the K's. If this element is an element of b, recompute the K's but not the V's.

v. Redo the estimation of the c's. Call the new estimate $\hat{c}^*$.

vi. Find the jth column of $\partial\hat{c}/\partial\hat{\alpha}$ as $(\hat{c}^* - \hat{c})/0.001$.

vii. Restore the jth element of $\hat{\alpha}$ to where it started.

viii. Go back to (iii), and repeat until you've done all the elements of α.

ix. Also compute the derivatives subtracting rather than adding 0.001 from each element of $\hat{\alpha}$. Each derivative should be close to the

corresponding one the first time around. If not, this will be a sign that our numerical derivatives may differ in an important way from the true derivatives.

B. COMPUTING THE STANDARD ERRORS WE WANT

We want to find the two-standard-error bands for the y_i^*'s. For a given i, the upper bound is the y such that $K_i - G_{iy}$ is twice its standard error; the lower bound is the y such that $G_{iy} - K_i$ is twice its standard error.

$K_i - G_{iy}$ takes the form $d'\omega$, so finding its standard error is conceptually straightforward. So the problem is mainly one of finding the upper and lower bounds for each i. What follows is my suggestion.

The upper bound:

Start at i = 1. The smallest possible y here is y = 9 (since a team can't get a first down inside its own 10). So compute $K_1 - G_{19}$ and its standard error. If $K_1 - G_{19}$ is more than twice its standard error, go on to i = 2, and compute $K_2 - G_{28}$ and its standard error, and see if $K_2 - G_{28}$ is more than twice its standard error. Proceed like this until you find a case where K - G for the smallest number of yards the team could have to go is less than twice its standard error. Then search over values of y in some way (for example, by incrementing in units of 0.01) until you find the y such that $K_i - G_{iy}$ is twice its standard error. This will give you y_i^{UPPER}. Don't go past the second decimal place in considering values of y.

Now set i = i + 1. Guess a y_i^{UPPER} (see the next paragraph) and compute $K_i - G_{iy}$ and its standard error. Then search over y's (for example, by incrementing or decrementing in units of 0.01) until you find y_i^{UPPER}.

For the initial guess, you could use $y_i^* + (y_{i-1}^{UPPER} - y_{i-1}^*)$. Near the team's own goal line, you won't have a y_{i-1}^*; in these cases, you could use $y = y_{i-1}^{UPPER}$.

Continue until you find an i such that y_i^{UPPER} exceeds the distance to the opponent's goal line.

The lower bound:

Start at i = 1. Compute $G_{iy} - K_i$ and its standard error for the smallest possible y (for i = 1, this is y = 9). Continue incrementing i until $G_{iy} - K_i$ for the smallest possible y (which is 0, not 1, for $i \geq 10$) is more than twice its standard error. Then search over y (for example, by incrementing in units of 0.01) until you find y_i^{LOWER}.

Now set i = i + 1, and proceed analogously to before. For an initial guess, I'd use $y = y_i^* - (y_{i-1}^* - y_{i-1}^{LOWER})$. If there isn't a y_{i-1}^{LOWER} (because $G_{i-1,0} - K_{i-1}$ was less than twice its standard error), I'd use an initial guess of $y_i^{LOWER} = 0$.

Proceed all the way to i = 99.

ANALYSIS - D.

2/20/02

III. WHAT SHOULD TEAMS DO?

1. For each i starting with one's own 9-yard line, compute the difference in the estimated value of going for it with 0.75 yards to go and the estimated value of kicking. This difference is $G_{i1} - K_i$.

a. Print out this difference for each i.

b. Plot the difference as a function of i.

2. Let's use y_i^* to denote the number of yards to go where the team is indifferent between kicking and going for it. That is, y_i^* is defined by $G_{iy_i^*} = K_i$. You can find the y_i^*'s either using the quadratic formula (but watch out for imaginary answers) or by some type of numerical search.

Plot y_i^* as a function of i.

(Also, it doesn't make sense to consider impossible situations: (i) If we find $y_i^* < 0$, such show it as 0 in the plot. (ii) If we find y_i^* exceeds the distance to the opponent's goal line, just set it to that distance in the plot. (iii) If we find y_i^* is less than the distance to one's 10-yard line, just set it to that distance in the plot.)

IV. WHAT DO TEAMS ACTUALLY DO, AND HOW MUCH COULD THEY GAIN BY CHANGING THEIR BEHAVIOR?

This uses your fourth-down spreadsheet.

A. HOW OFTEN DO TEAMS GO FOR IT WHEN THEY SHOULDN'T?

Find all fourth downs where $y > y_i^*$ for the team's position and where the team went for it ("G", perhaps with a number after it, in Column 6 of your fourth-down spreadsheet).

If there are only a few such cases, simply print out the row of the spreadsheet for each one.

If there are more than 25 or so, tell me, and we will discuss how to look at them.

Finally, find how many fourth downs there are where $y > y_i^*$ for the team's position and where the team kicked ("U" or "F", perhaps with a number after it, in Column 6 of your fourth-down spreadsheet).

B. HOW OFTEN DO TEAMS KICK WHEN THEY SHOULDN'T?

Let Max_i be the largest integer less than or equal to y_i^*. Exception: If $0.75 \le y_i^* < 1.00$, set $Max_i = 1$ rather than $Max_i = 0$.

Example 1: If $y_i^* = 3.17$, $Max_i = 3$. That is, for this i a team is better off

on average going for it if it has 1, 2, or 3 yards to go.

Example 2: If $y_i^* = 0.9$, $Max_i = 1$. That is, for this i a team is better off on average going for it if it is recorded as having 1 yard to go, since what's recorded by the NFL as 1 yard to go is on average 0.75 yards to go.

Construct a table with 99 rows. Each row should have Max_i entries (so some rows will have no entries). For example, suppose $Max_{70} = 2$ (on the opponent's 30, the team should go for it if it has 1 or 2 yards to go). Then row 70 should have 2 entries, the first for fourth and 1 on the opponent's 30, the second for fourth and 2 on the opponent's 30.

Each entry should consist of two numbers, ideally presented in the form A/B, where B is the number of times that fourth down arises (for example, fourth and 1 on the opponent's 30), and A is the number of times the team went for it on those fourth downs. (Don't forget that plays where the team has 1 yard to go appear as 0.75 yards to go in the spreadsheet.)

Finally, construct a second table with 99 rows; each row should again have Max_i entries. For row i, column y, report $G_{iy} - K_i$. (When doing the entries for column 1, report $G_{iy} - K_i$ with y = 0.75 rather than y = 1, for the usual reason.)

Also, report the sum over all the entries in the whole table of the A's and of the B's.

C. HOW LARGE ARE THE GAINS FROM DOING THE RIGHT THING?

For each entry in the table, find the number of times the team should have gone for it but didn't (in the notation above, this is B - A). Multiply this by the corresponding entry in the second table ($G_{iy} - K_i$). Sum over all entries in the table, divide by two, and then divide by the number of games (732). This is the average gain per team per game from better decisions about this issue.

TO BE CONTINUED ...

ANALYSIS - D.

4/23/02

1. The bottom-line calculations you gave me last week weren't quite right, because of how they treated the 98 and 99 yard lines. What I need are:

-- The sum over all the entries in the whole table of the A's and of the B's.

-- For each entry in the table, find the number of times the team should have gone for it but didn't (B - A). Multiply this by the corresponding entry in the second table (G_{iy} - K_i). Sum over all entries in the table.

ANALYSIS - D.

4/23/02

-- Specifics about computing 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. The only subtlety involves cases where we have no observations. In particular, if there is no y such that we have a positive number of observations with y or fewer yards to go and such that teams go for it at least as often as they kick when they have y or fewer yards to go, report 0 for that yard line. To make this clearer, here are some examples. The entries are in the usual form of [# of times going for it]/[# of cases].

Yd line	1 yd to go	2 yds	3 yds.	4 yds.	ANSWER
20	0/1	1/2	2/3	0/10	3
21	1/1	0/1	0/0	0/10	3
22	0/0	0/0	1/3	0/10	0 (not 2)
23	0/0	1/2	0/0	0/10	3
24	0/0	0/3	2/3	0/10	0 (not 1)
25	2/3	0/2	1/1	0/10	3

ANALYSIS - D.

12/6/04

I'd like to update the following statements from the paper:

(Note: Treat F2's, F4's, F5's, and U2's as G's. Treat F3's and U5's as kicks.)

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

"Consider teams that are trailing by at least 4 points in the third quarter. ... 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 kicked 270 times." (Note: As with the previous quote, exclude not only plain "O"'s, but also "O"'s with numbers and/or "*"'s after them.)

"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." (Note: Again, exclude not only plain "O"'s, but also "O"'s with numbers and/or "*"'s after them.)

"In the 732 first quarters in the sample, there are 992 cases where a team kicked when the difference between the estimated values of going for it and kicking was positive, or an average of 0.68 cases per team-quarter. The average estimated value of the expected gain from going for it in these cases is 0.37 points."

ANALYSIS - E.

3/6/02

The idea is to redo the earlier stuff you did about the relative values of field goals and touchdowns using data for 1998-2000. So let me just start by reprinting my old instructions (which where constructed mainly by quoting from the paper):

> 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 therefore overstates the value of a touchdown relative to a field goal, and thus overstates the benefits of going for it on fourth down.
>
> Three considerations suggest that this effect is not very important. [...]
>
> 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 one to overestimate 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 xxxx games in the sample 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 the 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 xxxx(xxxx-1)/2 (or xxxx'(xxxx'-1)/2) synthetic games for each decile, for a total of 2xxx 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. [Unless the game is tied, there is a 50-50 chance that the team is the trailing one. Thus if p{i} is the probability that a synthetic game has a margin of i points, the number we'd want for a field goal is .5*(p{0} + p{1} + p{2}) + .25*p{3}.] Although there are various reasons that this estimate is imperfect, it is likely to provide a reasonable

approximation to the value of an additional score in the first quarter.

The exercise suggests that 3 additional points the first quarter raise a team's probability of winning by xxxx percentage points, and that 7 additional points raise the probability by xxxx percentage points. Thus, [to be filled in]. [In addition to seeing the bottom line from this calculation, I'd like to see the full distribution of the p{i}'s.]

ANALYSIS - F.

12/6/04

V. ACTUAL DECISIONS OVER ALL FOUR QUARTERS

I want to look at teams' fourth-down decisions over the whole game, not just in the first quarter. For now, I just want summary statistics.

Times and scores to consider:

1. First quarter
2. All four quarters
3. Quarters 2, 3, and 4
4. Trailing by 4 or more, quarters 2, 3, and 4
5. Quarter 2, trailing by 4 or more; plus Quarter 3, trailing by $4 \le x \le 28$; plus Quarter 4, trailing by $4 \le x \le 16$
6. Leading by between 10 and -3 (that is, leading by 1-10, trailing by 1-2, or tied), quarters 2 and 3
7. Fourth quarter, trailing by at least 4 and no more than 14

Fourth down positions to consider:

1. All cases where the recommended choice is kicking
2. All cases where the recommended choice is going for it
3. All cases where the recommended choice is going for it and the team is in its own half of the field
4. All cases where the recommended choice is going for it and the team has 5 or more yards to go

What to report: For each of the 7x4 = 28 combinations, report how many fourth downs there were, and how many times the team went for it. As usual, exclude not only plain "O"'s, but also "O"'s with numbers and/or "*"'s after them. And as before, treat F2's, F4's, F5's, and U2's as G's, and treat F3's and U5's as kicks.

ANALYSIS - G.

9/17/02

I'd like to look at momentum issues.

All you'll need for this is the situations spreadsheet and the V's. The first step is to find places where good and bad things happen.

Good things: All first-down situations where, between the situation and the next one, the team with the ball scored a touchdown. That is, you need to find the first-down situations with 6.984 in the appropriate column of the spreadsheet. Call the situation before the touchdown A, the one after B, and the next 3 C, D, and E.

Bad things: All first-down situations where either: (i) the team that did not have the ball in the situation scored a touchdown before the next situation (so -6.984 in the relevant column); or (ii) there was no score before the next situation, and the team that did not have the ball in the situation has a first down no more than 10 yards downfield from the original situation. (Example: OAK is playing SF, and has first and 10 at its own 40. Then if there's no score, and the next situation is a first down for SF at the 50 or anywhere in OAK territory, this satisfies (ii).) Again use A, B, C, D, E, to denote the starting situation and the subsequent 4.

So we now have 2 samples (good things and bad things). Each observation within a sample consists of 5 situations. There will sometimes be overlap between observations. For example, if there is a turnover on one play (without forward movement of more than 10 yards) and a touchdown on the next, the initial situation will be the "A" for a bad observation, and the next situation will be both the "B" for that observation and the "A" for a good observation.

In some cases, we won't have the data for some or all of B, C, and D, because of the end of the first quarter. For those observations, just use what we have.

For each observation, compute the "gain" from A to B, B to C, C to D, and D to E, <u>always from the point of view of the team that had the ball in situation A</u>. To clarify what I mean by gain, consider B to C. By gain, I mean: net points scored between B and C plus value of situation C minus value of the situation B.

Now regress each of the 8 gains ({good,bad} x {A to B, B to C, C to D, D to E}) on a constant. Don't worry about correcting the s.e.s.

ANALYSIS - H.

12/6/04

I. FOURTH DOWNS VS. THIRD DOWNS

A. SAMPLES

Consider fourth-down plays coded as G's (including ones with numbers and/or "*"'s after them).

I want to consider 4 samples:

1. First quarter only.
2. First 3 quarters, excluding last two minutes of the first half.
3. Whole game, excluding last two minutes of each half and excluding overtime.
4. Whole game, excluding overtime.

B. VARIABLES

For each fourth-down play where the team went for it:

Let X be the realized value of the play (the net number of points scored before the next situation plus the value of the next situation from the perspective of the team that had the ball).

Let G be our estimate of the value of going for it from the yard line and yards to go the team had (from our analysis of third downs).

Let PS be the amount the team with the ball was favored by.

Let D the amount the with the ball was leading by.

C. REGRESSIONS

For each sample:

1. Compute the means of PS and D.

2. Run the following two regressions:

   X - G on a constant

   X - G on a constant, PS, and D

D. EXPLANATION

If third downs are a good proxy for fourth downs, the X's (the payoffs to going for it on fourth down) should on average equal the G's (the estimated value of going for it based on third-down data). Thus this hypothesis predicts that regressing X - G on a constant will yield a coefficient of zero.

The reason for looking at PS and D is that there may be a tendency for teams that are underdogs and/or trailing to go for it, which would tend to make X on average less than G just because the sample is skewed toward bad teams.

ANALYSIS - I.

4/27/02

e. At one point you did some summary statistics for me for third-down plays with various number of yards to go. Since we now lump yard lines 1-83 together, I want to redo those statistics for those yard lines. Here are more old instructions with the yard lines changed:

> Considering only observations between one's own 1 and the opponent's 17 (inclusive), I'd like to know, for each number of yards to go (1 yard to go, 2 yards to go, ..., up through 15): (1) the number of observations with that number of yards to go; (2) the number of those times that the team would have made it (that is, the number of those times when either the entry in Column 9 of your spreadsheet is a 1 or the entry in Column 8 is strictly positive); (3) the average number of yards the ball moved forward (from the perspective of the team with the ball). (For (3), whenever the entry in Column 8 of your spreadsheet is 0, you can use the information in Columns 9-10 to figure out where the ball ended up. When the entry in Column 8 is not 0, compute the number of yards the ball moved as the number of yards to the relevant goal line -- the opponent's goal line if the number of points scored is positive, the team's own goal line if the number of points scored is negative.)

ANALYSIS - I.

4/23/02

5. I want to check whether sample-selection issues are messing up the estimation. The issue here is that as we get further down the field, the observations are increasingly skewed toward better teams.

So: Redo the regression estimating the V's (the values of the different situations) with another right-hand-side variable included. The new variable is the amount the team with the ball is favored by. Do it in the usual way (knot points at the 9's, 17's, 33's, and 50). Don't worry about the standard errors. The V's are then the fitted values evaluated at AMOUNT FAVORED = 0.

Also, to make sure we have a clean comparison, redo the old estimates. (That is, estimate the V's with the usual knot points, no amount favored on the right-hand side.) Print out and plot the two sets of estimated V's.

The attached spreadsheet has the point-spread information.

ANALYSIS - I.

5/9/02

I want to redo the regression estimating the V's (the values of the different situations) with the sample restricted to the one-third of games with the smallest point spreads. So:

-- Find the one-third of games (= 244 games) with the smallest point spreads (in absolute value). The point spreads are in LISTALL.XLS, where I assume you have at this point. Choose some way of dealing with the fact that there is unlikely to be a break in the point-spread distribution at exactly one-third of the sample.

-- Reestimate the V's using only observations from these games.

-- Use our preferred specification: knot points at the 9's, 17's, 33's, and 50.

-- Don't compute standard errors.

-- Print out and plot these V's together with our usual ones (full sample, usual knot points, no amount favored on the right-hand side, etc.).