$$V_y := \left[\begin{array}{l}\text{Value of starting at yd. line } y\\ \text{expected diff b/t points scored by team with}\\ \text{ball and opponent from now until the end of the game with}\\ \infty \text{ time}\end{array}\right]$$

$gt = (g\text{: game index})(t\text{: index of play number during game } g)$

y = yd. line at which the play starts for the team with the ball $(1, 2, \ldots, 49, 50, o49, o48, \ldots, o1)$ and (kickoff 30, kickoff 20)

$$D_{gt,y} = \mathbb{1}\{\text{play } gt \text{ starts at yd. line } y\}.$$

$$P_{gt} = \text{Net points scored by team with ball on play } gt \text{ before next play } g(t+1)$$
$$= \#\text{ points scored by team during play } gt - \#\text{ points scored by opponent}$$

$$B_{gt} = \begin{cases} 1 & \text{if team with ball on play } gt \text{ also has ball at play } g(t+1) \\ -1 & \text{else} \end{cases}$$

Value of a 1^{st} down at yard line y (in the 1^{st} Quarter)

$$\boxed{\text{Value of play } gt = \sum_y D_{gt,y} V_y = \mathbb{E}\left[P_{gt} + B_{gt} \sum_y D_{g(t+1),y} V_y\right]}$$

ERROR $\quad \varepsilon_{gt} = \left[P_{gt} + B_{gt} \sum_y D_{g(t+1),y} V_y\right] - \mathbb{E}\left[P_{gt} + B_{gt} \sum_y D_{g(t+1),y} V_y\right]$

$$\Rightarrow \quad \sum_y D_{gt,y} V_y = \left[P_{gt} + B_{gt} \sum_y D_{g(t+1),y} V_y\right] - \varepsilon_{gt}$$

$$\Rightarrow \quad P_{gt} = \sum_y \underbrace{\left(D_{gt,y} - B_{gt} D_{g(t+1),y}\right)}_{:= X_{gt,y}} V_y + \varepsilon_{gt}$$

$$\Rightarrow \quad \boxed{P_{gt} = \sum_y X_{gt,y} V_y + \varepsilon_{gt}}$$

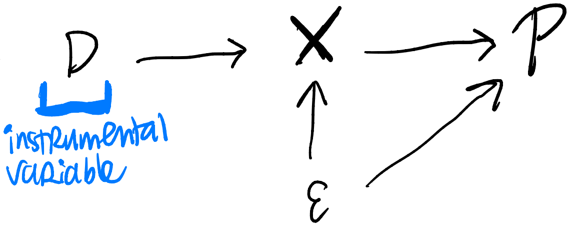
Regression?

$$\begin{bmatrix} P_1 \\ \vdots \\ P_n \end{bmatrix} = \begin{bmatrix} \quad X_{n \times p} \quad \end{bmatrix} \begin{bmatrix} V_1 \\ \vdots \\ V_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \qquad P = XV + \varepsilon$$

Correlation between residuals ε and design matrix X

Linear Model $\quad P = XV + \varepsilon$

Correlation Diagram



$$\boxed{\varepsilon_{gt} \text{ uncorrelated with } D_{gt,y} \quad \forall y}$$

$$\varepsilon_{gt} = \left[P_{gt} + B_{gt} \sum_y D_{g(t+1),y} V_y\right] - \mathbb{E}\left[P_{gt} + B_{gt} \sum_y D_{g(t+1),y} V_y\right]$$

- If ε_{gt} were correlated with some $D_{gt,y}$, then when a team is at yd. line y, the realized value of yd. line y one play later being at yd. line would be systematically different than the expected value of being at yd. line y, which is V_y, contradicting the def. of V_y

$$\boxed{\varepsilon_{gt} \text{ is possibly correlated with } X_{gt,y} \text{ via } -B_{gt} D_{g(t+1),y}}$$

$$X_{gt,y} = \left(D_{gt,y} - B_{gt} D_{g(t+1),y}\right)$$

Instrumental Variable Regression

- Simple Regression:

$$\begin{cases} P = \beta_0 + \beta_1 X + \varepsilon \\ \mathrm{cov}(D, \varepsilon) = 0 \text{ but } \mathrm{cov}(\varepsilon, X) \neq 0 \text{ and } \mathrm{cov}(D, X) \neq 0 \\ \to D \text{ is instrumental variable} \end{cases}$$

$$\Rightarrow \mathrm{cov}(P, D) = \beta_1 \mathrm{cov}(X, D) + \cancelto{0}{\mathrm{cov}(\varepsilon, D)}$$

$$\Rightarrow \beta_1 = \frac{\mathrm{cov}(P, D)}{\mathrm{cov}(X, D)}$$

- Extend to Multiple Linear Regression:

$$\boxed{\hat{\beta} = (D^T X)^{-1} D^T P}$$

$(X^T X)^{-1} X^T P$

since $$\hat{\beta} = \hat{\beta}_{P \text{ on } X} = \frac{\hat{\beta}_{P \text{ on } D}}{\hat{\beta}_{X \text{ on } D}} = \frac{(D^T D)^{-1} D^T P}{(D^T D)^{-1} D^T X} = (D^T X)^{-1} D^T P.$$

Quadratic Spline

101 covariates is a lot.

- V_y for all yd. lines $\{1, 2, \dots, 49, 50, {}_{o}49, {}_{o}48, \dots, {}_{o}1\} \cup \{\text{kickoff } 30, \text{ kickoff } 20\}$

value of 1st down <u>smooth function</u> of team's position on field

→ estimated V_y's become a $\boxed{\text{quadratic spline}}$ as a function of team's position on the field with knot points at both 9-, 17-, 33-, 50 yd. lines

+ V_y for kickoffs

→ $\boxed{12}$ parameters instead of 101

$$\begin{cases} V_1(y) = a_1 y^2 + b_1 y + c_1, & y \in [1, 9] \\ V_2(y) = a_2 y^2 + b_2 y + c_2, & y \in (9, 17] \\ V_3(y) = a_3 y^2 + b_3 y + c_3, & y \in (17, 33] \\ \vdots & \vdots \\ V_8(y) = a_8 y^2 + b_8 y + c_8, & y \in ({}_o9, {}_o1] \end{cases}$$

Constraints at knot points y_0

$$\begin{cases} V_k(y_0) = V_{k+1}(y_0) \\ V_k'(y_0) = V_{k+1}'(y_0) \end{cases}$$

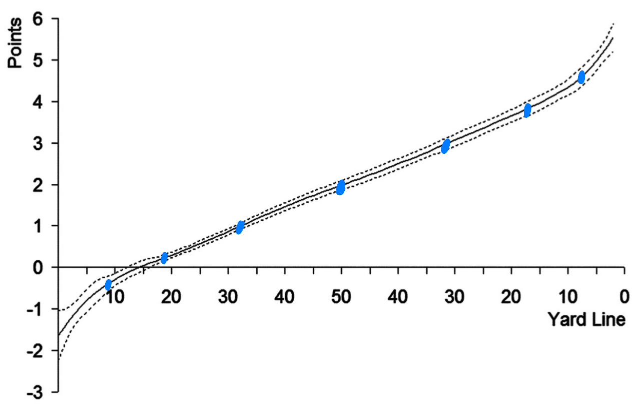
$$\begin{cases} a_k y_0^2 + b_k y_0 + c_k = a_{k+1} y_0^2 + b_{k+1} y_0 + c_{k+1} \\ 2a_k y_0 + b_k = 2a_{k+1} y_0 + b_{k+1} \end{cases}$$

↳ determines 2 params b_{k+1}, c_{k+1}

Why 12 params

3 params for $V_1(y)$
1 param for $V_2(y)$ (since $V_1(y) = V_2(y)$, $V_1'(y) = V_2'(y)$ determines 2 params of $V_2(y)$)
1 param for $V_3(y)$
1 param for $V_8(y)$
1 param for kickoff 30 (not quadratic spline)
1 param for kickoff 20 (not quadratic spline)

Results – Value of a 1st and 10 at given yd. line



Kicking vs. Going for It

Value of Kicking

Data All kicks in 1^{st} quarter of 1998-2000

(punts
field goal attempts
blocked & muffed kicks
kicks nullified by penalties)

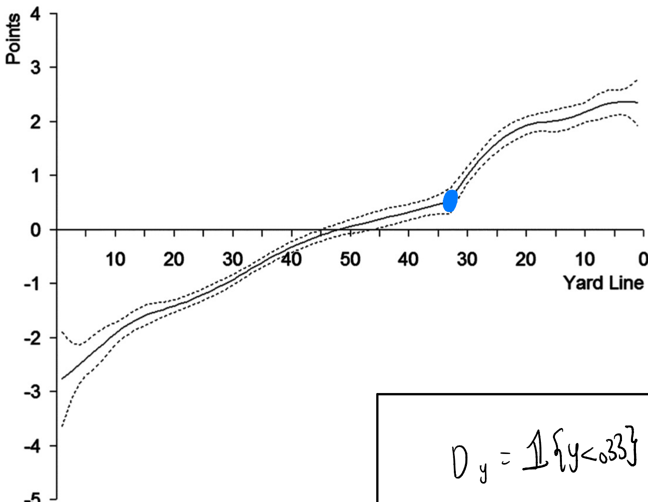
Method Avg. the realized value of the kicks from a given yd. line, including the net points scored on that play, and the value of the subsequent yd. line (using previous V_y results).
Constrain estimated values of kicks to be smooth, with 1 knot point at 33 yd. line (where the choice between punting and FGA changes).

quadratic spline again except not equal slope at knot point

$$
\begin{cases}
KV_1(y) = a_1 y^2 + b_1 y + c_1, & y < .33 \\
KV_2(y) = a_2 y^2 + b_2 y + c_2, & y \geq .33 \\
KV_1(.33) = KV_2(.33)
\end{cases}
$$

kick value

Estimated Value of Kicking as a function of team's position on the field



(a)

$D_y = \mathbb{1}\{y < .33\}$ still a regression

$$
\begin{bmatrix} | & | & | & | & | & | \\ D_y y^2 & D_y y & D_y & (1-D_y) y^2 & (1-D_y) y & (1-D_y) \\ | & | & | & | & | & | \end{bmatrix}
\begin{bmatrix} a_1 \\ b_1 \\ c_1 \\ a_2 \\ b_2 \\ c_2 \end{bmatrix}
=
\begin{bmatrix} 3 + V_{y'} \\ 3 + V_{y'} \\ 0 + V_{y'} \\ \vdots \end{bmatrix}
$$

further transform into spline...

Value of Going for It

all 3rd down plays in 1st quarter of 1998–2000

Data · Because teams rarely go for it on 4th down (in 1998-2000), use 3rd down plays instead

Method · Would like to, given the yd. line on the field and the number of yds. to hit 1st down, simply average all realizations. But there are 1000+ cases, so need to smooth the estimates.

· Instead, focus on the difference between going for it and of turning the ball over on the spot.

Difference between Values of Going for it and of turning the ball over on the spot

is $G_{yf} - (-V_{y'}) = G_{yf} + V_{y'}$ on yd. line y, with f yards to get to 1st down, where y' is "opposite" yd. line

going for it depends on yd. line y only thru the opponent's value if you turn it over! (for $y > 17$)

$y \in$ [team goal line, opponent 17] $\Rightarrow$ $G_{yf} + V_{y'} = a_0 + a_1 f + a_2 f^2$ → independent of y, quadratic in f

$y \in$ [opponent 17, opp goal line] $\rightarrow$ $G_{yf} + V_{y'} = b_0 + b_1 f + b_2 y + b_3 y^2 + b_4 fy + b_5 f^2 + b_6 f^2 y + b_7 f y^2 + b_8 f^2 y^2$ → quadratic in f and y

$y =$ opp. 17 $\Rightarrow$ function and derivative constrained to be equal

similarly, turn it into a spline regression!

Difference between values of Going for it and of turning the ball over on the spot

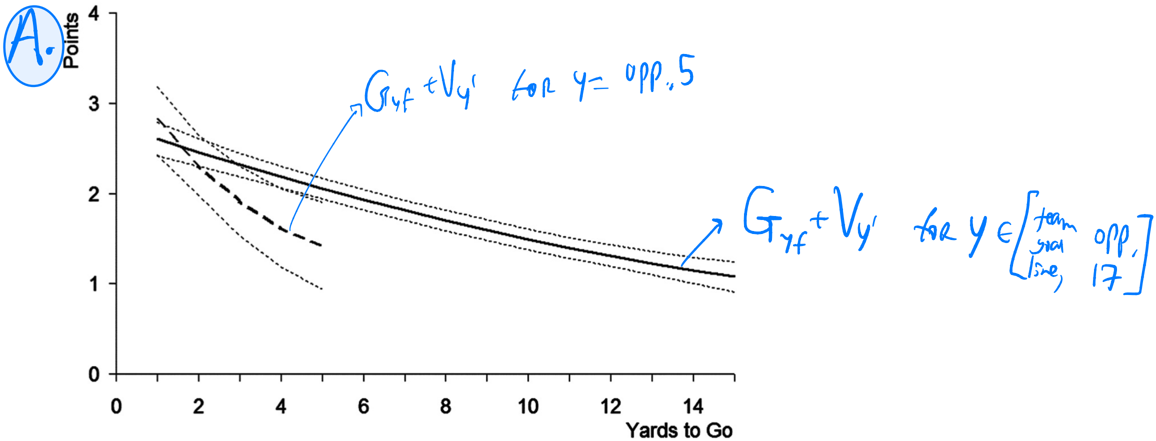
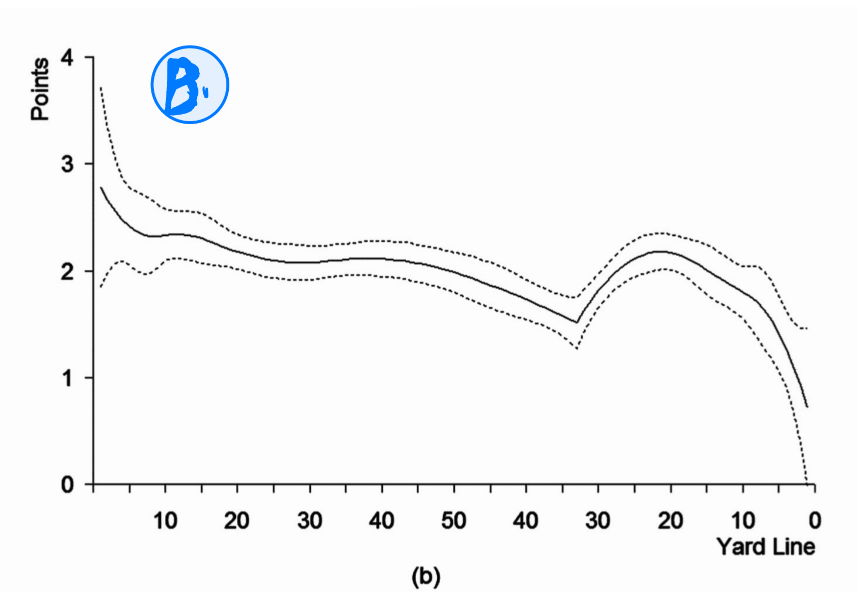
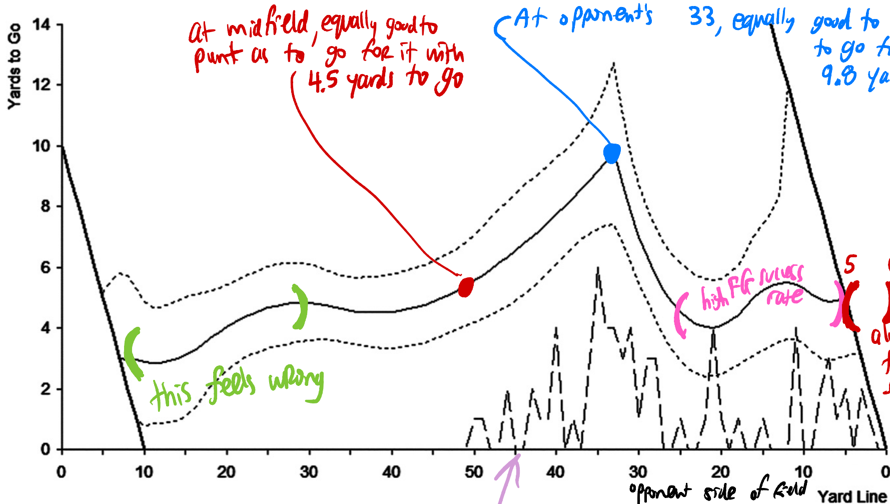


FIG. 3.—The estimated difference between the values of going for it and of the other team having the ball on the spot at a generic yard line outside the opponent's 17 (solid line) and at the opponent's 5 (dashed line). The dotted lines show the two-standard-error bands.

Difference between estimated value of a kick and of the other team having a 1st down on the spot



# of Yards to go where, given the yd. line,
Value of Kicking = Value of going for it



at midfield, equally good to punt as to go for it with 4.5 yards to go

At opponent's 33, equally good to kick as to go for it with 9.8 yards to go!!

this feels wrong

high FG success rate

5 0

always better to go for it, same FG success rate but higher value from TD rate & bad field position if fail

opponent side of field

FIG. 4.—The number of yards to go where the estimated values of kicking and going for it are equal (solid line) and two-standard-error bands (dotted lines), and the greatest number of yards to go such that when teams have that many yards to go or less, they go for it at least as often as they kick (dashed line).

Get this plot by comparing plots A. and B.

dashed line is completely under solid line, means teams' cutoff yd. line for going for it rather than kicking is lower than it should be