Statistical Methods III: Spring 2013

Jonathan Wand

Stanford University

Choice models: generalize III

Outline

Spatial model

Consider three parties in unidimensional model

- where should parties locate?
  depends on whether there is threat of entry or not? stackleberg equilibrium
  most stat models treat location as given
- what party should a party vote for?
- how should voter make choice?

Axiomatic Foundations of Choice Models

Assumptions

- D1 Let $R \subset S \subset T \subset U$.
- D2 Let $x, y, z \in T$, arbitrary elements of choice set.
- D3 Let $P(x, y)$ be the probability of choosing x instead of y, $0 < P(x, y) < 1$.
- D4 $P_S(R)$ is the probability of choosing R given choice from among alternatives in S.

Choice Axiom

- (i) $P_T(R) = P_S(R)P_T(S)$
- (ii) If $P(x, y) = 0$ for some $x, y \in T$, $P_T(S) = P_{T-\{x\}}(S - \{x\})$

Axiomatic Foundations of Choice Models

Axiom of Choice

- Defines relationship which defines how choices within subsets are related in the context of an individual making probabilistic choices.
- Can be rewritten as $P_T(R \mid S)P_T(S) = P_T(R)$
- Two core implications,
  Lemma 3: Independence of Irrelevant Alternatives (IIA)
  Theorem 3: Probability must satisfy a ratio scale

Axiomatic Foundations of Choice Models

Lemma 3 (Independence from irrelevant alternatives):
For $x, y \in S$,

$$\frac{P(x,y)}{P(y,x)} = \frac{P_S(x)}{P_S(y)}$$

Proof:
By Axiom we have

$$P_S(x) = P(x,y)[P_S(x) + P_S(y)]$$

So

$$\begin{aligned}
P_S(x) &= P(x,y)[P_S(x) + P_S(y)] \\
P_S(x) &= P(x,y)P_S(x) + P(x,y)P_S(y) \\
(1 - P(x,y))P_S(x) &= P(x,y)P_S(y) \\
P(y,x)P_S(x) &= P(x,y)P_S(y) \\
\frac{P(x,y)}{P(y,x)} &= \frac{P_S(x)}{P_S(y)}
\end{aligned}$$

Axiomatic Foundations of Choice Models

Luce Lemma 3 (Independence from irrelevant alternatives):
What does this mean?

- relative probability of choosing two alternatives is invariant to the composition of the larger set of alternatives.
- Only ratio is invariant, not probabilities themselves
- Might also hear that log-odds of two choices are constant: $\log(P_S(x)) - \log(P_S(y)) = c$.

Axiomatic Foundations of Choice Models

Luce Lemma 3 (Independence from irrelevant alternatives):
Why so cool?

- can estimate parameters defining utility of choices even with only a subset.
- ** Not generally possible if IIA does not hold (e.g., correlation between utilities of choices)—then to estimate any choice must model all choices.
- ** Neither holds in general for models of choice (by design) nor is it plausible that it in general holds empirically.

Axiomatic Foundations of Choice Models

Theorem 3: choice probability is ratio scale
$\exists v : T \to \Re_+$, unique up to multiplication by $k > 0$, such that

$$P_S(x) = \frac{v(x)}{\sum_{y \in S} v(y)} = \frac{1}{1 + \sum_{y \in S - \{x\}} v(y)/v(x)}$$

McFadden and Yellot have each pointed out the connection to logit models by setting $v(x) = e^x$.

E.g.,

$$P(x, y) = \frac{v(x)}{v(x) + v(y)} = \frac{e^{\mu_x}}{e^{\mu_x} + e^{\mu_y}} = \frac{1}{1 + e^{\mu_y}/e^{\mu_x}} = \frac{1}{1 + e^{-(\mu_x - \mu_y)}}$$

Yellot shows that discriminal process based on Type I discrete value distribution is uniquely equivalent to Choice Axiom.

Axiomatic Foundations of Choice Models

Let $S \in \{1, 2, 3\}$, and $P_S(j)$ be probability of choosing j from S,

$$
\begin{aligned}
P_S(y=1) &= \frac{1}{1 + e^{\mu_2-\mu_1} + e^{\mu_3-\mu_1}} = \frac{e^{\mu_1}}{e^{\mu_1} + e^{\mu_2} + e^{\mu_3}} \\
P_S(y=2) &= \frac{1}{1 + e^{\mu_1-\mu_2} + e^{\mu_3-\mu_2}} = \frac{e^{\mu_2}}{e^{\mu_1} + e^{\mu_2} + e^{\mu_3}} \\
P_S(y=3) &= \frac{1}{1 + e^{\mu_1-\mu_3} + e^{\mu_2-\mu_3}} = \frac{e^{\mu_3}}{e^{\mu_1} + e^{\mu_2} + e^{\mu_3}}
\end{aligned}
$$

So,

$$
\frac{P_S(y=1)}{P_S(y=2)} = \frac{e^{\mu_1}}{e^{\mu_2}}
$$

Axiomatic Foundations of Choice Models

Let $T \in \{1, 2\}$, and $P_T(j)$ be probability of choosing j from T,
Recal logit (special case of MNL),

$$P_T(y = 1) = \frac{1}{1 + e^{\mu_2 - \mu_1}} = \frac{e^{\mu_1}}{e^{\mu_1} + e^{\mu_2}}$$
$$P_T(y = 2) = \frac{1}{1 + e^{\mu_1 - \mu_2}} = \frac{e^{\mu_2}}{e^{\mu_2} + e^{\mu_2}}$$

So,

$$\frac{P_T(y = 1)}{P_T(y = 2)} = \frac{e^{\mu_1}}{e^{\mu_2}}$$

Axiomatic Foundations of Choice Models

Comparing probabilities of choosing 1 and 2 in logit and MNL,

$$\frac{P_T(1)}{P_T(2)} = \frac{e^{\mu_1}}{e^{\mu_2}} = \frac{P_S(1)}{P_S(2)}$$

MNL conforms to Choice Axiom/IIA.

See Yellot (1977) and McFadden (1973) for connections between Luce and EV Type 1.

Classical experiment

Let (Y, X, D) be observable random variables,

1. Y_i is the outcome of interest
2. X_i a “pre-determined” variable
3. D_i indicator for treatment status

Researcher chooses an assignment mechanism. E.g.,

1. randomly draw a unit from a population
2. assign treatment $D_i = 1$ to unit with probability p

Some notation – potential outcomes

For dichomtous treatments,

- $Z_i \in \{0, 1\}$: assigment indicator
- $D_i \in \{0, 1\}$: received treatment indicator

allowing possibility that $Z_i \neq D_i$
Define potential outcomes, $Y_i(D_i)$:

- $Y_i(1)$: outcome if treated
- $Y_i(0)$: outcome if not treated

later we will also consider $Y_i(Z_i, D_i)$

- $Y(0, 0)$: neither assigned nor treated
- $Y(0, 1)$: not assigned, yet treated
- $Y(1, 0)$: assigned yet not treated
- $Y(1, 1)$: assigned and treated

Fundamental Problem of Causal Inference

Causal Effect (CE),

$$\tau_i = Y_i(1) - Y_i(0)$$

Key ideas:

- cannot observe both $Y_i(1)$ and $Y_i(0)$.
- causal effect may be heterogeneous (no structure in definition) hence τ_i is indexed by i

What do we observe?

$$\begin{aligned} Y_i &= D_i Y_i(D_i = 1) + [1 - D_i] Y_i(D_i = 0) \\ &= Y_i(D_i = 0) + D_i[Y_i(D_i = 1) - Y_i(D_i = 0)] \\ &= Y_i(D_i = 0) + D_i \tau_i \end{aligned}$$

which is control outome, plus CE if treated

Examples of Estimands

We can't observe τ_i, could we estimate the following?
Population average treatment effect, PATE

$$\tau_P = E[Y_i(1) - Y_i(0)]$$

Population average treatment effect for treated, PATT

$$\tau_{P,T} = E[Y_i(1) - Y_i(0) \mid D_i = 1]$$

Notes:

- Hm, still involve both $Y_i(1)$ and $Y_i(0)$ for each i...
- often will refer to PATE simply as ATE.

Difference in observed outcomes

Consider

$$\frac{1}{n_t} \sum_{i:D_i=1} Y_i(1) - \frac{1}{n_c} \sum_{i:D_i=0} Y_i(0)$$

with some regularity conditions, as sample of each treatment gets big...

$$\text{plim}\ \frac{1}{n_t} \sum_{i:D_i=1} Y_i(1) \to E[Y_i(1) \mid D_i = 1]$$

$$\text{plim}\ \frac{1}{n_c} \sum_{i:D_i=0} Y_i(0) \to E[Y_i(0) \mid D_i = 0]$$

Question: how does the limiting conditional difference,

$$E[Y_i(1) \mid D_i = 1] - E[Y_i(0) \mid D_i = 0] \tag{1}$$

relate to PATE

$$\tau_P = E[Y_i(1)] - E[Y_i(0)]$$

Does (1) identify a quantify of interest?

A point worth emphasizing

Coherent to refer to $Y_i(0)$ for someone assigned to $D_i = 1$?
or to refer to $Y_i(1)$ for someone assigned to $D_i = 0$?

	Unobserved			Perfect Doctor		
Unit	Truth		Unit CE	Assigned	Observed	
i	$Y_i(0)$	$Y_i(1)$	$Y_i(1) - Y_i(0)$	D_i	$Y_i(0)$	$Y_i(1)$
1	13	14	1	1	?	14
2	6	0	-6	0	6	?
3	4	1	-3	0	4	?
...						
6	6	1	-5	0	6	?
7	8	10	2	1	?	10
8	8	9	1	1	?	9

Conditional expectation $E[Y_i(j)|D = j]$ depends on assignement mechanism.

Comparing treated and control

$$
\begin{aligned}
& E[Y_i \mid D_i = 1] - E[Y_i \mid D_i = 0] \\
& = E[Y_i(1) \mid D_i = 1] - E[Y_i(0) \mid D_i = 0] \\
& = E[Y_i(1) \mid D_i = 1] - E[Y_i(0) \mid D_i = 0] + E[Y_i(0) \mid D_i = 1] - E[Y_i(0) \mid D_i = 1] \\
& = \{E[Y_i(1) \mid D_i = 1] - E[Y_i(0) \mid D_i = 1]\} + \{E[Y_i(0) \mid D_i = 1] - E[Y_i(0) \mid D_i = 0]\} \\
& = E[Y_i(1) - Y_i(0) \mid D_i = 1] + \{E[Y_i(0) \mid D_i = 1] - E[Y_i(0) \mid D_i = 0]\}
\end{aligned}
$$

Describe each of these in words

- $E[Y_i(1) \mid D_i = 1] - E[Y_i(0) \mid D_i = 0]$: avg difference in outcomes
- $E[Y_i(1) - Y_i(0) \mid D_i = 1]$: average treatment effect on treated
- $E[Y_i(0) \mid D_i = 1] - E[Y_i(0) \mid D_i = 0] \neq 0$: if selection bias

Question: what could you do, knowing this decomposition?

Difference in observed outcomes - randomized

What does simple randomization give us?

$$Y_i(0), Y_i(1) \perp\!\!\!\perp D_i$$

therefore,

$$E[Y_i(0) \mid D_i = 0] = E[Y_i(0) \mid D_i = 1] = E[Y_i(0)]$$
$$E[Y_i(1) \mid D_i = 0] = E[Y_i(1) \mid D_i = 1] = E[Y_i(1)]$$

Thus, no selection bias,

$$0 = E[Y_i(0) \mid D_i = 1] - E[Y_i(0) \mid D_i = 0]$$

and thus,

$$E[Y_i \mid D_i = 1] - E[Y_i \mid D_i = 0]$$
$$= E[Y_i(1) - Y_i(0) \mid D_i = 1] + \{E[Y_i(0) \mid D_i = 1] - E[Y_i(0) \mid D_i = 0]\}$$
$$= E[Y_i(1) - Y_i(0) \mid D_i = 1] = \text{PATT}$$
$$= E[Y_i(1) - Y_i(0)] = \text{PATE}$$

Let's generalize to accomodate differences conditional on pre-treatment variables...

- allow for different assignement probabiliteis as a function of observables
- what do we need to assume to identify ATE in this more general case?

Key concepts: Unconfoundedness

An assignment is unconfounded if the assignment mechanism is independent of the potential outcomes.

$$Y_i(0), Y_i(1) \perp\!\!\!\perp D_i \mid X_i$$

(can also be unconditional)

Many labels for essentially the same idea:

- "ignorable treatment assignment"
  Rosenbaum and Rubin (1983)
- "conditional independence assumption"
  Lechner (1999, 2002)
- "selection on observables"
  Barnow, Cain, and Goldberger (1980)

Key concepts: Unconfoundedness

In experiments,

- may stratify on an observable variable (or more than one)
- stratification useful to ensure both treatments within a rare group

In observational studies,

- applications approximating an experiment seek to justify unconfoundeness is acheived prior to making comparisons
- assumes that all variables necessary for understanding assignment mechanism are observable

Key concepts: Unconfoundedness – violations

Perfect Pollster

$$D_i = \begin{cases} 1 & \text{if } Y_i(1) > Y_i(0) \\ 0 & \text{if } Y_i(1) \leq Y_i(0) \end{cases}$$

Pretty good pollster, $p > .5$

$$P(D_i = 1) = \begin{cases} p & \text{if } Y_i(1) > Y_i(0) \\ 1 - p & \text{if } Y_i(1) \leq Y_i(0) \end{cases}$$

Key concepts: Probabilistic assignment

Probabilistic assignment, for all X,

$$0 < Pr(D_i = 1 \mid X) < 1$$

Also known as

- "overlap"
- common support on X

Key concepts: Probabilistic assignment – violation

If $Pr(D_i = 1 \mid X) = 1$
then there exists no observation of $E[Y \mid X, D = 0]$.
(and mirror problem for $Pr(D_i = 1 \mid X) = 0$)

For example,

X_i	$E[Y_i(0) \mid x_i]$	$E[Y_i(1) \mid x_i]$	$P(D_i = 1 \mid x_i)$
0	0	1	0.4
1	2	5	1

Question:

- can we still make inference for subset where $X = 0$?

An alternative estimand - conditional

Conditional average treatment effect, CATE

$$\tau_C = \frac{1}{N} \sum_i E[Y_i(1) - Y_i(0) \mid X_i]$$

Conditional average treatment effect for treated, CATT

$$\tau_{C,T} = \frac{1}{N_T} \sum_{i:D_i=1} E[Y_i(1) - Y_i(0) \mid X_i]$$

Notes:

- ATE conditional on sample distribution of X
- intrinsically of interest if representativeness of sample is dubious
- and can condition on subset of X (it is pre-treatment)
- in matching methods, will often trim to area of common support.

Unconfoundness + Overlap = (C)ATE identifiable

By unconfoundedness,

$$E[Y \mid D = d, X = x] = E[Y(d) \mid D = d, X = x]$$
$$= E[Y(d) \mid X = x]$$

By overlap, we observe for each subpoluation x,

$$E[Y \mid D = 1, X = x] - E[Y \mid D = 0, X = x]$$
$$= E[Y(1) \mid D = 1, X = x] - E[Y(0) \mid D = 0, X = x]$$
$$= E[Y(1) \mid X = x] - E[Y(0) \mid X = x]$$
$$= E[Y(1) - Y(0) \mid X = x]$$
$$= \tau(x)$$

Can weight over distribution of X in sample to get CATE, or population distribution of X to get ATE, $E[\tau_i]$

Endogenous assignement

$$C_i = 1 \text{ if } a + bX_i + U_i > 0, \text{ else } C_i = 0. \quad (1)$$

In application, $C_i = 1$ means that subject i self-selects into treatment. The second equation defines the subject's response to treatment:

$$Y_i = 1 \text{ if } c + dZ_i + eC_i + V_i > 0, \text{ else } Y_i = 0. \quad (2)$$

Endogenous assignement

Step 1. Estimate the probit model (1) by likelihood techniques.

Step 2. To estimate (2), fit the expanded probit model

$$P(Y_i = 1|X_i, Z_i, C_i) = \Phi(c + dZ_i + eC_i + fM_i) \tag{3}$$

to the data, where

$$M_i = C_i \frac{\phi(a + bX_i)}{\Phi(a + bX_i)} - (1 - C_i) \frac{\phi(a + bX_i)}{1 - \Phi(a + bX_i)}. \tag{4}$$

Review: Bivariate normal

- Take $X, Z \sim N(0,1)$, independent
- Set $Y = \rho X \; + \; \sqrt{1-\rho^2}\, Z$
- Note: $E(X) = E(Y) = 0$,
  $Var(X) = Var(Y) = 1,\ Corr(X,Y) = \rho$

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix} \cdot \begin{bmatrix} X \\ Z \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

- $\Rightarrow$ (X, Y) has the "standard bivariate normal" distribution

Endogenous assignement

Consider (1–2). We can represent V_i as $\rho U_i + \sqrt{1-\rho^2} W_i$, where W_i is an $N(0, 1)$ random variable, independent of U_i. Then

$$\begin{aligned} E\left\{V_i \middle| X_i = x, C_i = 1\right\} &= E\left\{\rho U_i + \sqrt{1-\rho^2} W_i \middle| U_i > -a - bx_i\right\} \\ &= \rho E\{U_i | U_i > -a - bx_i\} \\ &= \rho \frac{1}{\Phi(a + bx_i)} \int_{-a-bx_i}^{\infty} x\phi(x)\mathrm{d}x \\ &= \rho \frac{\phi(a + bx_i)}{\Phi(a + bx_i)} \end{aligned} \tag{9}$$

because $P\{U_i > -a - bx_i\} = P\{U_i < a + bx_i\} = \Phi(a + bx_i)$. Likewise,

$$E\left\{V_i \middle| X_i = x, C_i = 0\right\} = -\rho \frac{\phi(a + bx_i)}{1 - \Phi(a + bx_i)}. \tag{10}$$

Endogenous assignement

Step 1. Estimate the probit model (1) by likelihood techniques.

Step 2. To estimate (2), fit the expanded probit model

$$P(Y_i = 1|X_i, Z_i, C_i) = \Phi(c + dZ_i + eC_i + fM_i) \tag{3}$$

to the data, where

$$M_i = C_i \frac{\phi(a + bX_i)}{\Phi(a + bX_i)} - (1 - C_i) \frac{\phi(a + bX_i)}{1 - \Phi(a + bX_i)}. \tag{4}$$

Endogenous assignement

Table 1 Simulation results

	c	d	e	ρ
True values				
	−1.0000	0.7500	0.5000	0.6000
Raw estimates				
Mean	−1.5901	0.7234	1.3285	
SD	0.1184	0.0587	0.1276	
Two-step				
Mean	−1.1118	0.8265	0.5432	
SD	0.1581	0.0622	0.2081	
MLE				
Mean	−0.9964	0.7542	0.4964	0.6025
SD	0.161	0.0546	0.1899	0.0900

Notes. Correcting endogeneity bias when the response is binary probit. There are 500 repetitions. The sample size is 1000. The correlation between latents is $\rho = 0.60$. The parameters in the selection equation (1) are set at $a = 0.50$ and $b = 1$. The parameters in the response equation (2) are set at $c = -1$, $d = 0.75$, and $e = 0.50$. The response equation includes the endogenous dummy C_i defined by (1). The correlation between the exogenous regressors is 0.40. MLE computed by VGAM 0.7-6.

Endogenous assignement

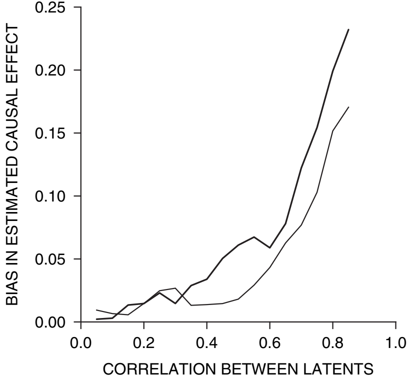

Fig. 1 The two-step correction. Graph of bias in $\hat{e}$ against ρ, the correlation between the latents. The light lower line sets the correlation between regressors to 0.40, and the heavy upper line sets the correlation to 0.60. Other parameters as for Table 1. Below 0.35, the lines crisscross.

Londgrean et al 1995

Londregan, Bienen; and van de Walle. 1995. Ethnicity and Leadership Succession in Africa. *ISQ*.

Question: What is the effect of leader's own ethnic population share on non-constitutional replacement of leader?

Shows larger share increase probability of non-constitutional replacement (but often replaced from within own ethnic group).

Key measure is (E)thnic (S)ize (D)ominance.

Thus we create a measure which adjusts the ethnic share of the leader's group for the degree of diffusion among the country's ethnic groups. We call this measure "ESD_1" (ethnic size dominance). This measure accounts for both size and dispersion of ethnic groups, and is derived from what is called a Herfindahl index. (See Herfindahl, 1950; Stigler, 1968; Hart, 1971.) Our ESD_1 measure for leader L is defined as follows:

$$ESD_1 = \frac{S_L}{\sqrt{S_1^2 + S_2^2 + \ldots + S_N^2}}$$

Case study: Londregan et al

TABLE 2. Descriptive Statistics

Variable	*Mean*	*Std. Dev.*	*Min.*	*Median*	*Maximum*	*% Within*
Annual growth rate	0.01	0.07	−0.40	0.01	0.44	92.27
ln(income)	6.76	0.56	5.40	6.79	8.40	10.28
Openness to trade	0.34	0.22	0.04	0.28	1.42	28.33
Leader's ethnic share	0.36	0.27	0.01	0.29	0.99	10.21
ESD_1	0.59	0.30	0.02	0.63	0.99	27.42
ESD_2	0.29	0.29	0.00	0.18	0.99	8.43
Political exit	0.08	0.27	0.00	0.00	1.00	91.94
Nonconstitutional exit	0.07	0.25	0.00	0.00	1.00	91.63
Nonconstitutional entrant	0.40	0.49	0.00	0.00	1.00	40.82
Inter-ethnic leadership transition	0.05	0.21	0.00	0.00	1.00	93.40

Case study: Londregan et al

TABLE 3. Sample Correlations

	Log of Lagged Income	*Openness to Trade*	*Ethnic Herf. Index*	*Leader's Ethnic Share*	*ESD_1*	*ESD_2*	*Political Exit*
Income growth rate	−0.06	−0.01	−0.06	0.10	0.06	0.11	−0.14
	(−1.64)[a]	(−0.19)	(−1.53)	(2.99)	(1.67)	(3.19)	(−3.87)
Log of lagged income		0.37	−0.20	0.26	0.22	0.27	0.00[b]
		(13.59)	(−5.21)	(8.52)	(7.10)	(8.99)	(−0.12)
Openness to trade			−0.02	0.11	0.06	0.14	−0.08
			(−0.44)	(3.34)	(1.82)	(4.31)	(−2.27)
Ethnic Herfindahl index				−0.72	−0.43	−0.70	0.02
				(−15.76)	(−10.44)	(−15.47)	(0.45)
Leader's ethnic share					0.89	0.99	−0.05
					(78.70)	(372.52)	(−1.37)
ESD_1						0.88	−0.03
						(71.43)	(−0.71)
ESD_2							−0.05
							(−1.33)

[a]T–ratios in parentheses.
[b]The estimated correlation between the log of lagged real per capita income and our political exit variable is −0.004, which is 0.00 to two decimal places.

Case study: Londregan et al

TABLE 5. Ethnicity and Nonconstitutional Succession

Dependent Variable: Nonconstitutional Exit				
Variable	*(1)*	*(2)*	*(3)*	*(4)*
Previous year's log of per capita income	−2.47	−2.53	−3.21	−3.73
	(1.15)[a]	(4.46)	(1.28)	(1.49)
Nonconstitutional ruler	2.95	2.41	3.10	2.88
	(1.13)	(1.37)	(1.24)	(1.29)
ln(leader's ethnic share)			5.46	
			(3.17)	
ln(leader's ethnic share)^2			1.70	
			(1.02)	
ln(ESD_1)				8.70
				(4.06)
ln(ESD_1)^2				5.37
				(2.69)
Log likelihood function[b]	−22.24	−6.61	−19.79	−18.02
Sample size	67	67	67	67

[a]Standard errors in parentheses.

[b]In column (2) this is the log of the conditional likelihood function corresponding to the conditional logit, and is not comparable with the logit likelihoods reported in columns (1), (3), and (4).

Case study: Londregan et al

The net impact of our ESD_1 variable is a weighted average of the coefficients of the log of ESD_1 and its square. More precisely, the estimated impact is given by:

$$\frac{p'(\underline{x}'\underline{\hat{\beta}})\left\{\hat{\beta}_1 + 2\hat{\beta}_2 \ln(ESD_1)\right\}}{ESD_1}$$

Estimates of this effect are shown in Figure 2. These estimates are calculated for a probability of nonconstitutional transition at the mean for the subsample of exit observations: 0.82. There we see that for low values of ESD_1 the estimated impact of an increase in ESD_1 on the probability a leadership transition is nonconstitutional is negative. However, these estimated negative effects are statistically insignificant, as indicated by the wide confidence bands, which encompass 0. At higher levels of ESD_1, the effect reverses: for values above 0.57, increases in ESD_1 significantly increase the probability that leadership transitions take place by nonconstitutional means. This critical value is just below the sample median of 0.63, so that, for just over half of our sample, the effects of increasing the relative size of the leader's ethnic group are directly counter to the prediction of the hypothesis

Case study: Londregan et al

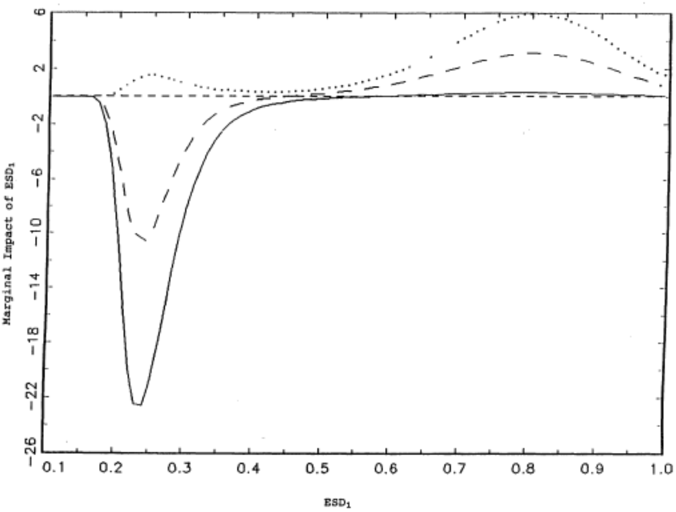

FIG. 2. ESD_1 vs. its marginal impact on the probability a leadership transition takes place nonconstitutionally.