Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 10176–10180, December 1989
Political Sciences

Aggregate-level analysis and prediction of midterm senatorial elections in the United States, 1974–1986

ALLAN J. LICHTMAN* AND V. I. KEILIS-BOROK†

*Department of History, The American University, Washington, DC 20016; and †Institute of the Theory of Earthquake Prediction and Mathematical Geophysics, Academy of Sciences of the U.S.S.R., U.S.S.R. B. Grusikskaya 10, Moscow 12810, U.S.S.R.

Contributed by V. I. Keilis-Borok, September 11, 1987; revision received January 12, 1989

ABSTRACT **Pattern recognition study demonstrates that the outcomes of American midterm senatorial elections follow the dynamics of simple integral parameters that depict preelectoral situations aggregated to the state as a whole. A set of "commonsense" parameters is identified that is sufficient to predict such elections state-by-state and year-by-year. The analysis rejects many similar commonsense parameters. The existence and nature of integral collective behavior in U.S. elections at the level of the individual states is investigated. Implications for understanding the American electoral process are discussed.**

In a previous paper (1) we presented a synthetic approach to studying the American electoral process, verifying the hypothesis that the outcome of presidential elections follows the dynamics of simple integral parameters that depict social, economic, and political conditions aggregated to the nation as a whole. In the present paper, we test the hypothesis that integral parameters diagnose the outcomes of elections even on a subnational level, i.e., for senatorial elections in each of the individual American states.

As in ref. 1, we use the pattern recognition of small samples, a method that has been successful in many fields where qualitative factors have to be analyzed and the volume of data is insufficient for the application of more traditional statistical procedures. These fields include earthquake prediction, medical diagnosis, geological exploration, speech recognition, and, in political science, the prediction of judicial decisions (2).

As a recent review of the literature indicates (3), political studies have generally focused on the outcomes of congressional rather than senatorial elections. In fact, as shown in a summary of forecasts for the 1986 senatorial contests (4), political science models do not predict the winners and losers in individual states, but only the nationwide division of seats between Republicans and Democrats. In 1986 as in past years, individual analysts presented a range of contradictory forecasts for senatorial elections, some of which were highly accurate. But their procedures were neither reproducible nor necessarily applicable to future contests (4, 5).

In this paper we attempt to design for midterm senatorial elections the year-to-year and state-by-state prediction system that thus far has eluded students of senatorial elections. Our purpose is to find among the many potentially relevant parameters a set that is sufficient for prediction and provides stable results. We use the same simple version of pattern recognition that we previously applied to the study of presidential elections (ref. 1; updated and corrected tables on presidential elections are available from the authors).

DATA AND ALGORITHM

We analyzed 99 American midterm senatorial elections held in 1974, 1978, and 1982. Elections, identified by the state and year, are divided into two classes: *I*, the incumbent party retains the Senate seat (whether or not the incumbent senator was a candidate for reelection), and *C*, the challenging party captures the Senate seat.

Table 1. Questionnaire for midterm Senate elections: An answer of "yes" favors the incumbent Senate party

1. The incumbent-party candidate is the sitting senator.
2. The incumbent-party candidate is a major national figure.
3. There was no serious contest for the incumbent-party nomination.
4. The incumbent party won the seat with 60% or more of the vote in the previous election.
5. The challenging-party candidate is not a national figure or a past or present governor or member of Congress.
6. There was no serious contest for the challenging-party nomination.
7. The incumbent-party candidate is not of the same party as the President.
8. The incumbent-party candidate outspends the challenger by 10% or more.

The situation before each election in each state is described by a set of discretized ("yes" or "no") answers to eight questions, formulated in Table 1. The questions are the same for each state and election, but the answers vary according to the circumstances prevailing before an upcoming election.

Several of the questions in Table 1 require clarification. Question 2: a major national figure is an individual who would be generally known by the educated public outside the home state. Examples among current senators would include E. Kennedy, J. Glenn, and R. Dole. Question 3: there is a serious contest for the incumbent-party nomination when the winning candidate receives less than two-thirds of a primary election vote. Question 6: there is a serious contest for the challenging-party nomination when the winning candidate fails to gain a majority of the vote *and* at least twice the vote cast for the next highest finisher. Finally, answers to the finances question (number 8) require projections from the preelection financial reports submitted to the Federal Election Commission and other available information.

Only sufficiency, but neither optimality nor completeness, is claimed for this set of questions. Other questions may be equally or even more relevant to diagnosing the outcome of Senate elections. If, however, our questionnaire were sufficient to predict with high reliability the results of upcoming elections, it would disclose empirical regularities in Senate elections that would both guide and restrict our understanding of how such elections actually work.

Table 2 displays the answers to all eight questions for each election. This constitutes our learning material. As in ref. 1 our problem is to derive from it a "rule of recognition" that can assign an election to class *I* or *C* with minimal chance of an error, given the answers to the questionnaire for that particular contest.

Using the Hamming algorithm of pattern recognition (1) we formed a "kernel" representing the answers that favored

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Downloaded by guest on December 4, 2021

Table 2. Responses to questionnaire

Incumbent-party victories																		Challenger victories								
State/	Answer to question*								State/	Answer to question*								State/	Answer to question*							
year	1	2	3	4	5	6	7	8	year	1	2	3	4	5	6	7	8	year	1	2	3	4	5	6	7	8
AK74	0	1	1	1	0	1	0	0	NM78	0	1	0	1	0	0	0	0	CO74	0	1	0	1	1	1	1	0
AL74	0	1	0	0	0	0	0	0	OR78	0	0	0	1	0	0	0	0	FL74	1	1	1	1	0	1	1	1
AR74	1	1	1	1	0	0	0	0	RI78	0	1	0	1	0	0	1	0	KY74	0	1	0	1	1	0	1	1
AZ74	0	0	0	1	0	0	1	0	SC78	0	0	0	0	0	0	0	0	NH74	1	1	0	1	0	1	1	1
CA74	0	1	0	1	0	0	0	0	TN78	0	0	0	0	0	1	0	0	NV74	1	1	1	1	1	0	0	1
CT74	0	0	0	1	0	0	0	0	TX78	0	0	0	1	1	1	0	0	OH74	1	1	1	1	0	1	1	1
GA74	0	0	0	0	0	0	0	0	VA78	1	1	0	1	0	0	0	0	VT74	1	1	1	0	0	0	1	1
HI74	0	0	0	0	0	0	0	0	WV78	0	1	0	0	1	0	1	0	IA78	0	1	0	1	0	1	1	0
IA74	0	1	0	1	0	0	0	0	WY78	1	1	1	0	1	1	0	0	CO78	0	1	0	1	1	0	1	1
ID74	0	0	0	0	0	0	0	0	AZ82	0	1	0	1	0	1	0	0	MA78	0	1	0	1	0	1	0	0
IL74	0	0	0	1	0	0	0	0	CA82	1	1	1	1	1	0	1	0	ME78	0	1	0	1	1	0	1	1
IN74	0	0	0	1	0	0	0	0	CT82	0	0	0	1	1	0	1	0	MI78	0	1	0	1	1	1	0	0
KS74	0	1	0	0	0	0	1	0	DE82	0	1	0	1	0	0	1	0	MN78	1	1	1	1	0	0	1	1
LA74	0	0	0	0	0	0	0	0	FL82	0	1	0	0	0	1	0	0	MS78	1	1	1	1	1	0	1	1
MD74	0	1	0	1	0	1	1	0	HI82	0	1	0	1	0	1	0	0	NE78	1	1	0	1	1	0	0	1
MO74	0	0	0	1	0	0	0	0	IN82	0	0	0	1	1	1	1	0	NH78	0	1	0	1	0	1	1	1
NC74	1	1	1	0	0	0	0	0	MA82	0	1	0	1	0	0	0	1	NJ78	1	1	1	0	1	0	0	1
ND74	0	1	0	0	1	0	1	0	MD82	0	1	0	1	0	0	0	0	OK78	1	1	0	1	1	1	0	1
NY74	0	0	0	1	0	1	1	0	ME82	0	1	0	0	1	0	0	0	SD78	1	1	1	1	1	0	1	1
OK74	0	1	0	1	0	1	1	0	MI82	0	1	0	1	0	1	0	0	NM82	0	1	0	1	0	1	1	1
OR74	0	1	0	1	0	1	1	0	MN82	1	1	0	0	0	0	1	1	NV82	0	1	1	0	0	1	0	1
PA74	0	1	0	1	0	1	1	0	MO82	0	1	0	1	0	1	1	0									
SC74	0	1	0	0	0	0	0	0	MS82	0	0	0	0	0	0	0	1									
SD74	0	0	0	1	0	1	0	0	MT82	0	1	0	0	0	0	0	0									
UT74	1	1	0	1	1	0	1	1	ND82	0	1	0	0	0	0	0	0									
WA74	0	1	0	0	0	0	0	0	NE82	0	1	0	1	0	1	0	1									
WI74	0	1	0	0	0	0	0	0	NJ82	1	1	1	0	0	1	0	0									
AK78	0	1	0	0	0	1	0	0	NY82	0	0	0	1	0	1	0	0									
AL78	1	1	1	0	0	0	1	0	OH82	0	1	0	1	0	0	0	0									
AR78	1	1	1	0	0	0	1	0	PA82	0	1	0	1	0	0	1	0									
DE78	0	1	0	1	0	1	1	0	RI82	0	1	0	1	0	0	1	0									
GA78	0	1	0	1	0	1	1	0	TN82	0	1	0	1	1	0	0	0									
ID78	0	1	0	1	0	0	0	0	TX82	0	0	0	1	1	1	0	0									
IL78	0	0	0	0	0	0	0	0	UT82	0	1	0	1	0	0	1	0									
KS78	1	1	1	0	0	0	0	1	VT82	0	1	1	1	0	0	1	0									
KY78	0	1	0	1	0	1	1	0	WA82	0	0	0	0	0	1	0	0									
LA78	0	1	0	1	0	0	1	0	WI82	0	0	0	0	0	0	0	1									
MT78	1	1	0	1	0	0	1	0	WV82	0	0	0	0	1	0	0	0									
NC78	0	0	0	1	0	1	0	0	WY82	0	1	0	1	0	0	1	0									

Incumbent parties were identified according to results of the previous election, including two cases with appointees of the opposition party.
*0, Yes; 1, no.

victory by an incumbent-party candidate. Each question listed in Table 1 is phrased so that the preferential answer is "yes." Using these answers we determined for each election the Hamming distance D from the kernel, that is, the number of "no" answers.

0	1	2	3	4	5	6	7
			WY82I				
			UT82I	VT82I			
			TX82I	NJ82I			
			TN82I	NE82I			
			RI82I	MO82I			
			PA82I	IN82I			
	WV82I		MN82I	MT78I			
	WI82I		MI82I	KY78I			
	WA82I	OH82I	MA82I	KS78I			
	ND82I	NY82I	HI82I	GA78I			
	MT82I	ME82I	DE82I	DE78I			
	MS82I	MD82I	CT82I	AR78I			
	TN78I	FL82I	AZ82I	AL78I	WY78I		
	OR78I	NM78I	WV78I	PA74I	NM82C		
	WI74I	NC78I	VA78I	OR74I	NJ78C		
	WA74I	ID78I	TX78I	OK74I	NH78C		
SC78I	SC74I	AK78I	RI78I	MD74I	NE78C	CA82I	
IL78I	MO74I	SD74I	LA78I	AR74I	ME78C	UT74I	
LA74I	IN74I	KS74I	NC74I	AK74I	CO78C	OK78C	SD78C
ID74I	IL74I	IA74I	NY74I	MI78C	VT74C	MN78C	MS78C
HI74I	CT74I	CA74I	ND74I	NV82C	KY74C	NV74C	OH74C
GA74I	AL74I	AZ74I	MA78C	IA78C	CO74C	NH74C	FL74C

FIG. 1. Division of off-year elections (1974–1982) by the Hammings distance (D) from the kernel (i.e., by the number of answers "no" to the questionnaire in Table 1). Each election is represented by the two-letter state abbreviation, the year, and the outcome: I, incumbent-party victory; C, challenging-party victory.

Fig. 1 shows the values of D for each election studied. From 1974 to 1982, incumbent-party victories have been characterized by $D < 5$ in 75 out of 78 cases, and incumbent-party losses by $D \geq 5$ in 17 out of 21 cases. This in turn suggests the following rule of recognition for elections I and C: classify an election as I, if $D < 5$; as C, if $D \geq 5$.

The eight questions used in this way to distinguish incumbent- from challenging-party victories were selected from a wider set, which included also the questions listed in Table 3. The selected questions divided past elections into I and C with the fewest errors.

To test our rule of recognition on independent data, we applied the rule prospectively to the midterm senatorial elections of 1986. The results of this test are described in the next section.

PREDICTION AND ITS STATISTICAL SIGNIFICANCE

Prediction. Table 4 reports the value of D for each senatorial election held in 1986 and the corresponding prediction

Downloaded by guest on December 4, 2021

Table 3. Questions considered, but not included in the questionnaire

1. Was the previous Senate election won by the incumbent party?
2. Was the incumbent party's Senate seat won by a margin of 55% or more?
3. Did the incumbent party win four or more of the past seven Senate elections?
4. Is the other Senate seat held by the incumbent party?
5. Was the Senate seat held by the incumbent party for two or more terms?
6. Is the state unemployment rate for the election year less than the national rate?
7. Is the state unemployment rate for the election year equal to or lower than the rate for the previous year?
8. Did the state unemployment rate increase less or decrease more during the election year than did the national rate?
9. Is the state per capita income equal to or higher than the national per capita income for the election year?
10. Did the state per capita income increase at a rate equal to or greater than the national increase during the election year?
11. Did the state per capita income increase at a rate equal to or greater than the national increase during the past 6 years?
12. Is the ADA (Americans for Democratic Action) rating of the incumbent senator significantly different from that of the state's senators and representatives combined?
13. Did the vote for the incumbent party in the previous presidential election equal or exceed the national mean for that party?
14. Did a Senate seat change parties in four or more of the state's previous seven elections?
15. Did the incumbent Senate party win the previous presidential election?

as published 1 week *prior* to the elections (5). Also included in Table 4 are the actual results of each election and the *Congressional Quarterly*'s qualitative classifications of the prospects for incumbent parties in the 1986 elections (6). Our predictions were correct for 30 elections out of 34.

We test the statistical significance of our predictions in order to demonstrate that the questionnaire casts light on the process responsible for the outcomes of Senate elections (purpose A) and is useful for predicting the results of future contests (purpose B).

For purpose A, it is sufficient to demonstrate that the sum of the answers to our questionnaire correlates with the outcomes of all 34 elections. That these results can be predicted independently by our set of parameters singles this set out among the numerous factors purportedly related to the mechanism of electoral decisions.

For purpose B, we also need to test the statistical significance of our predictions only for those elections with outcomes considered to be uncertain by authorities in the field. For a rigorous test, we eliminated from the score of successes those contests that authorities qualified as relatively secure for one or the other party (6).

Competing Hypothesis. The statistical significance of our results is tested by estimating the probability of obtaining equal or greater success by random prediction, based on the following competing hypothesis: each incumbent belonging to the presidential party wins a senatorial election with some level of probability P_p and loses with probability $1 - P_p$; each incumbent not belonging to this party wins with probability P_o and loses with probability $1 - P_o$; the results of elections for different states are independent (actually, these probabilities may be different for various groups of states; this is considered below).

We characterize the quality of prediction by the value T, the total number of correct predictions. Let us denote the following: $F_T(x)$, the probability distribution function of the value T under the assumption that the competing hypothesis described above is correct; T_a, the actual number of correct predictions made by our recognition rule (Table 4); ε, the probability to get $\geq T_a$ successes by chance under the competing hypothesis. Then

$$\varepsilon = 1 - F_T(T_a). \qquad [1]$$

Accordingly, the confidence level of the competing hypothesis (that our recognition rule provides nonrandom prediction) is $1 - \varepsilon$.

To evaluate ε we count the number of 1986 elections falling into each of four categories: n_{pp}, n_{po}, n_{op}, and n_{oo}, where the first index indicates whether the incumbent belonged to the presidential (p) or opposing (o) party, and the second index indicates the party affiliation of the winner. We then count the number of elections in each category correctly predicted by our recognition rule. These counts are denoted by ν_{iw} (each index i and w can be equal either to p or to o); the first index again indicates the incumbent party, and the second indicates the winner.

According to the competing hypothesis, the values ν_{iw} have

Table 4. Predictions and actual results for the 1986 election

	Answer to question*											
State	1	2	3	4	5	6	7	8	*D*	CQ† (ref. 6)	Prediction (ref. 5)	Actual result
HI	0	0	0	0	0	1	0	0	1	P.S.	Win	Win
OH	0	0	0	0	1	0	0	0	1	P.S.	Win	Win
SC	0	0	0	0	0	1	0	0	1	P.S.	Win	Win
UT	0	0	0	0	0	0	1	0	1	P.S.	Win	Win
AK	0	1	0	1	0	0	0	0	2	P.S.	Win	Win
CT	0	1	0	1	0	0	0	0	2	P.S.	Win	Win
KS	0	0	0	0	0	1	1	0	2	P.S.	Win	Win
KY	0	1	0	0	0	1	0	0	2	P.S.	Win	Win
ND	0	1	0	0	0	0	1	0	2	P.V.	Win	Loss‡
AR	0	1	0	1	0	0	1	0	3	P.S.	Win	Win
CA	0	0	0	1	1	1	0	0	3	V.	Win	Win
IL	0	1	0	1	0	1	0	0	3	P.S.	Win	Win
IN	0	0	1	1	0	0	1	0	3	P.S.	Win	Win
IA	0	1	0	1	0	0	1	0	3	P.S.	Win	Win
NH	0	1	0	1	0	0	1	0	3	P.S.	Win	Win
OR	0	0	1	1	0	0	1	0	3	P.V.	Win	Win
VT	0	1	0	1	1	0	0	0	3	P.V.	Win	Win
AZ	1	1	0	1	0	0	1	0	4	P.S.	Win	Win
CO	1	1	0	1	1	0	0	0	4	V.	Win	Win
ID	0	1	0	1	1	0	1	0	4	V.	Win	Win
LA	1	1	0	0	1	0	0	1	4	H.V.	Win	Win
NY	0	1	0	1	0	1	1	0	4	P.S.	Win	Win
NC	0	1	0	1	1	0	1	0	4	P.V.	Win	Loss‡
OK	0	1	0	1	1	0	1	0	4	V.	Win	Win
WA	0	1	0	1	1	0	1	0	4	P.V.	Win	Loss‡
WI	0	1	0	1	0	1	1	0	4	V.	Win	Win
AL	0	1	0	1	1	1	1	0	5	V.	Loss	Loss
FL	0	1	0	1	1	0	1	1	5	H.V.	Loss	Loss
GA	0	1	0	1	1	1	1	0	5	P.V.	Loss	Loss
MO	1	1	0	1	1	0	0	1	5	H.V.	Loss	Loss
PA	0	1	0	1	1	1	1	0	5	V.	Loss	Win‡
MD	1	1	0	0	1	1	1	1	6	H.V.	Loss	Loss
NV	1	1	0	1	1	0	1	1	6	H.V.	Loss	Loss
SD	0	1	1	1	1	0	1	1	6	H.V.	Loss	Loss

*0, Yes; 1, no.

†*Congressional Quarterly* assessment: P.S., probably secure; P.V., potentially vulnerable; V., vulnerable; H.V., highly vulnerable.

‡Incorrect prediction.

Downloaded by guest on December 4, 2021

binomial distributions and function F_T is their convolution:

$$F_T(.) = B(./n_{pp}, P_p) * B(./n_{po}, 1 - P_p) * B(./n_{op}, 1 - P_o) * B(./n_{oo}, P_o) \ldots \quad [2]$$

Here B is the binomial distribution, the first parameter (n_{iw}) is the number of trials, the second parameter is the probability of success in a single trial, the points stand for the arguments of distributions B, and $*$ represents convolution.

This estimation may be biased in favor of our hypothesis, since the statistic T does not distinguish between correctly predicting incumbent victories and defeats, although the probability of predicting defeat by chance is much smaller than the probability of predicting victory by chance. To reduce this bias we also characterize the quality of prediction by another measure: T_c, the number of correct predictions of incumbent-party defeats (i.e., of elections C only); $T_c = \nu_{po} + \nu_{op}$. Accordingly, in Eq. **1** we replace F_T by the distribution function of T_c, which is

$$F_c(.) = B(./n_{po}, 1 - P_p) * B(./n_{op}, 1 - P_o). \quad [3]$$

Confidence Levels. For purpose A, we estimate the probability of obtaining by chance an equal or greater number of successful predictions for all 34 elections. For purpose B, we estimate the same probability after eliminating elections generally regarded as relatively certain victories for either party (highly vulnerable and probably secure categories in ref. 6).

The confidence level $(1 - \varepsilon)$ for each test is estimated by $I - 3$; the necessary values of n_{iw}, P_p, and P_o are calculated from the data in Table 2 and in ref. 5; the values of ν_{iw} are obtained from Table 4.

For all 34 elections of 1986 we have $T = 30$ and

$$n_{pp} = 13 \qquad \nu_{pp} = 12$$

$$n_{po} = 9 \qquad \nu_{po} = 6$$

$$n_{op} = 1 \qquad \nu_{op} = 1$$

$$n_{oo} = 11 \qquad \nu_{oo} = 11.$$

For off-year elections of 1974–1982 in all states we estimate $P_p = 0.674$ and $P_o = 0.874$. From Eq. **2** we have then $\varepsilon < 0.03\%$, which is sufficiently small to reject the competing hypothesis of random prediction. This conclusion is stable: for $T = 29$, $\varepsilon < 0.2\%$, and even for $T = 27$, $\varepsilon < 3\%$; with one more success, $T = 31$, $\varepsilon = 0.004\%$.

From Eq. **3**, with $T_c = 7$, we have $\varepsilon < 5\%$.

We also tested whether samples of elections more similar to those held in 1986 might give more adequate estimations of probabilities P_p and P_o, possibly reducing our confidence level. This test showed that for any reasonable sample of elections, the values P_p and P_o remain about the same, leaving ε sufficiently small to reject the random hypothesis. We considered the following samples: (*i*) elections held, as in 1986, during a Republican administration, in 1974 and 1982; (*ii*) midterm elections in the same 34 states as in 1986; (*iii*) midterm elections held in those 34 states during a Republican administration. The last two samples included all off-year elections after World War II, to avoid unreliably small sampling. The variation in values P_p and P_o remained within 0.1; $\varepsilon = 0.03\%$ is the largest value we found. It corresponds to all elections for 1974 and 1982 and $T = 30$; for $T = 29$ it would still be <2%. So, we may conclude, on an acceptable confidence level of about 99.97%, that the answers to our questionnaire are highly correlated with the outcomes of midterm senatorial elections.

To test our questionnaire on elections that authorities in the field deemed uncertain, we rely on ref. 6, in which all the senatorial elections of 1986 were divided into four groups, as shown in the table below.

Prospects for the incumbent party in 1986 (ref. 6)	No.	Actual outcomes for incumbent party: Won R	Won D	Lost R	Lost D	Average D
Highly vulnerable	6	0	1	4	1	5.3
Vulnerable	7	4	2	1	0	4.1
Potentially vulnerable	6	1	1	4	0	3.5
Probably secure	15	8	7	0	0	2.3

Here, R refers to Republican candidates and D to Democrats.

After eliminating the "probably secure" group (where all incumbents actually won), we have in the remaining 19 states $T = 15$ and

$$n_{pp} = 5 \qquad \nu_{pp} = 3$$

$$n_{po} = 9 \qquad \nu_{po} = 7$$

$$n_{op} = 1 \qquad \nu_{op} = 1$$

$$n_{oo} = 4 \qquad \nu_{oo} = 4.$$

The probabilities P_p and P_o were again estimated for different samples of elections in order to check the stability of ε. We considered (*i*) all off-year elections in these 19 states since World War II and (*ii*) the subset of sample *i* held during a Republican administration. For these samples the probabilities are virtually identical, lying within intervals $P_p = 0.69 \pm 0.3$; $P_o = 0.84 \pm 0.3$.

Analysis of our 1986 sample of uncertain elections for $T = 15$ gives $\varepsilon < 0.7\%$; and for $T = 14$ we would have $\varepsilon < 3\%$; statistic $T_c = 7$ gives the same estimate as previously, since we eliminated only elections I.

Thus, at a high confidence level our results are correlated with the outcomes of elections identified as difficult to predict.

DISCUSSION

Stability. The division of elections into classes I and C is not stable near the threshold distance $D = 5$; its change to 6 or especially to 4 would create many additional errors, as Table 4 shows. With only eight parameters this instability is hardly avoidable; it remains when parameters are counted in D with different weights.

That our prediction was successful despite this instability implies that each parameter is an essential indicator of electoral outcomes. It may also suggest that the parameters are mutually dependent so that a change in one is likely to be accompanied by changes in others. The elimination of one or more parameters leads to additional errors, even when using the weighted procedure.

Algorithm CORA-3 (ref. 7) was applied in an attempt to obtain a more comprehensive recognition rule, based on combinations of two or three parameters. However, the combinations so far have no clear meaning and the recognition of incumbent- and challenging-party victories is neither more accurate nor more stable than in Hamming's algorithm, a conclusion that did not change after incorporating the initially rejected questions from Table 3.

Downloaded by guest on December 4, 2021

By "simulation history," we illustrate the stability of our recognition rule through electoral history by the same experiment as in ref. 1. Replicating the analysis from the perspective of an observer prior to the 1978 and 1982 elections, we used all questions from Tables 1 and 3 to select a questionnaire, derive a recognition rule, and "predict" the outcomes of the forthcoming election. Using learning material from only the elections of 1974, we selected all the questions but the fourth from Table 1. For the "next" election, 1978, this kernel gives four errors. Using results from both the 1974 and 1978 elections, we would select the same eight questions as in Table 1, obviously giving the same results for 1982 as are reported in Fig. 1.

Discretization of Parameters. We eliminated certain information by the discretization of parameters to the lowest level of resolution, yes or no. Likewise, we sought to predict the winners of elections, but not their percentages of the vote. Similar "robust" methods are widely used not only in pattern recognition, but in many other branches of exploratory data analysis (8). The apparent loss of information involved has often produced the stable results that elude more "detailed" analyses that are subject to fluctuations in the values of particular variables.

The prediction of voting percentages, of course, would be quite a different problem, not considered here.

CONCLUSIONS: PERSPECTIVES ON SENATORIAL ELECTIONS

Collective Behavior. The finding that aggregate-level parameters can reliably anticipate the outcome of both presidential and senatorial elections points to an underlying regularity in collective behavior that transcends the clustering of voters into interest and attitudinal groups—traditionally the focal point of electoral study. Electoral behavior in the United States is highly integrated not only for the nation as a whole but also within and between the diverse American states. In all states, for each election year, the outcomes of elections for the Senate follow the same rules, no matter by how much the play of each game may change.

Sufficient Factors. The reduction of the initial 30 questions to the 8 actually chosen provides unexpected insight into the electoral process. Surprisingly, the selected questions do not measure directly many factors traditionally believed to be associated with the outcome of senatorial elections. These include (see Table 3) the ideologies of competing candidates, the partisan division of the state's electorate, the issues raised in campaigns, the past tendency of states to reject or reelect incombent parties, and campaign strategy and tactics.

Most surprising was the finding that Senate incumbents do not suffer from a bad economy or benefit from a good one, no matter whether the economic measure is for the nation as a whole, for the individual state, or a comparison between the two. Incumbent-party candidates fared worst in the strong economies of 1978 and 1986, losing 12 and 10 seats, respectively, compared with losses of 7 seats in the 1974 recession and only 2 in the deep recession of 1982. Rather than punishing the party holding a Senate seat for hard times, some voters may instead regard the incumbent party as a safe port in a storm.

Comparisons Between Presidential and Senatorial Elections. There are notable similarities and differences in the regularities that underlie senatorial and presidential elections (1). Like the presidential elections, midterm Senate contests do not respond to factors such as party affiliation, issues, ideology, or the conduct of campaigns. But the particular characteristics of candidates—their political experience and stature—are much more important in Senate contests than in presidential elections. Whereas presidential elections turn on only a single, extraordinary characteristic—whether a candidate is charismatic or a national hero—Senate contests turn on such lower-threshold characteristics as whether either candidate is a national figure and whether a challenger has served as a governor or a member of Congress.

The greatest difference between the senatorial and presidential models is that the senatorial parameters include no specific performance indicators, suggesting that voters have more diffuse expectations of senators than of presidents. For Senate elections, perceptions of incumbent performance are captured only indirectly through questions about nomination contests, the political stature of opposition-party candidates, and the competition for financial support.

The Regularities of Complex Systems. That a single set of simple, integral parameters anticipates the results of a complicated electoral process is a familiar finding of the integral analysis of large, chaotic systems, especially in the physical sciences. Often, system-level regularities cannot be inferred from the study of micro-level interactions alone. Electoral systems may thus possess features similar to the large-scale physical systems that likewise exhibit collective behavior comprehensible only at the level of the system as a whole.

1. Lichtman, A. J. & Keilis-Borok, V. I. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 7230–7234.
2. Kort, F. (1957) *Am. Political Sci. Rev.* **51,** 1–12.
3. Wright, G. C., Jr., & Berkman, M. B. (1986) *Am. Political Sci. Rev.* **80,** 567–588.
4. Lewis-Beck, M. (1987) *Public Opinion*, 57–58.
5. DeCell, K. (1986) *The Washingtonian*, Nov., 142–145.
6. U.S. Congress (1986) *Congressional Q. Weekly Rep.*, Feb. 22, 1986, 336.
7. Gelfand, I. M., Guberman, Sh. A., Keilis-Borok, V. I., Knopoff, L., Press, F., Ranzman, E. Ya., Rotwain, I. M. & Sadovsky, A. M. (1976) *Phys. Earth Planet. Inter.* **11,** 227–283.
8. Tukey, J. (1977) *Exploratory Data Analysis* (Addison-Wesley, Reading, MA).

Downloaded by guest on December 4, 2021