Appendixes

APPENDIX A

Media and Demographic Measures from Chapter 3

Media Market Measures

"Area of dominant influence" (ADI) is a concept developed by Arbitron, Inc., to categorize every county in the United States as a television market. An ADI is determined via viewership surveys and television broadcast ranges. County:ADI mapping is exhaustive and exclusive. *TV Households* refers to the Arbitron estimate of households in counties (and larger aggregates) that have televisions. *Population* refers to the census person counts for these same statistical units. Congressional districts are designated CD. *Political district* is meant to refer to both states and CDs.

Variable Construction

Television market structure is measured in three ways: contiguity or "news mix", market dominance, and cost. Each taps a different dimension of media market efficiency, which affects the strategic decisions of campaigners.

News mix (Stewart and Reynolds 1990; Stewart 1989b) or *congruence* (Campbell, Alford, and Henry 1984) refers to the degree of overlap between a political district and the various television markets that serve it. Three distinct patterns of overlap have to be accounted for: a small television market encompassed by a large political district, a small political district encompassed by a large television market, and a district that is contiguous with only part of a television market or is covered by two or more markets.

News mix is a product of the proportion of the television market that is in the political setting and the proportion that is in the media market, summed across all relevant media markets. Suppose that:

- $\text{ADI}_1 - \text{ADI}_n$ = areas of dominant influence in a district, expressed as population size,
- P_D = the population of the political district,
- $\text{ADI}_n \cap P_D$ = the intersection of ADI_n and district D

Using these assumptions, contiguity is:

$$NM_D = \sum_{i=1}^{n} \frac{\text{ADI}_n \cap P_D}{P_D} * \frac{\text{ADI}_n \cap P_D}{\text{ADI}_n}$$

The measure runs from 1.0 (perfect overlap) to nearly zero.

Why call this measure "contiguity"? Suppose the proportion of coverage a television station allocates to a political campaign is a function of 1) the proportion of the ADI that is in the political district and 2) the proportion of the political district that is in the ADI. The product of these is used as a rough measure of the proportion of news that will be devoted to a particular district's political campaigns.

California's Nineteenth Congressional District is an example of nearly perfect overlap. The boundaries of the Monterey ADI and the CD are nearly identical. Delaware and Wyoming are examples of very inefficient media markets (among states with Senate races, they have the least-contiguous media markets). Delaware is a small part of a large ADI (Philadelphia), while Wyoming is served by six separate ADIs. Nevada is a nice illustration of a state that has a congressional district (the Second) with efficient media markets and another district (the First) with an only moderately efficient media market. As a state, Nevada has relatively efficient media markets. The Las Vegas ADI is completely contained within the state's borders, and the Reno ADI serves only 61,343 citizens not residing in Nevada. The main drag on the efficiency measure is caused by the 25,436 citizens who are served by the Salt Lake City market. These citizens are a small proportion of the ADI (lowering the efficiency measure), but they are also a small proportion of the total state (thus lowering their weight in the overall calculation). Looking at the Second District, 17 percent of its population is in the Las Vegas ADI, and only 15 percent of the *ADI* population is in the Second District. This is counterbalanced by the close fit between the Reno ADI and the rest of the Second District: 76 percent of the district's population lives in the Reno ADI, and the Reno district makes up 82 percent of the total ADI. While I may not expect that a great deal of coverage was devoted to the Second District race by Las Vegas television stations, the other 76 percent of the district could be expected to have received a heavy dose of campaign news. The result is a moderately high score on contiguity for this CD relative to other CDs. Finally, Wyoming is an example of a highly inefficient media market. If a candidate were to be so unwise as to blanket the state (i.e., to buy media time such that he or she appeared in every possible media market), that candidate would have to buy in six different markets with a population total of 4,565,409 (only 457,380 of whom live in Wyoming!). This is not a cost effective way to reach potential voters.

News mix takes into account all the ADIs in a political district. *Dominance* focuses on one part of this calculation; it shows whether the political district is dominated by a few ADIs (Stewart and Reynolds 1989). Dominance for district D (DOM_D) squares the first part of the contiguity calculation (proportion of the district in the ADI), in order to give greater weight to large overlaps, and then the squares across ADIs. DOM_D also runs from 1.0 (identical political district and ADI boundaries) to near zero:

$$\mathrm{DOM}_D = \sum_{i=1}^{n} \frac{\mathrm{ADI}_n \cap P_D}{P_D}.$$

Nearly all CDs have low contiguity scores, as would be expected (congressional districts seldom make up the bulk of a television market, and in many cases [New York City, Los Angeles] they make up quite a small proportion). The obverse side of the coin is that almost all are dominated by a single market. *If* he or she has the money, it is easy for a congressional candidate to decide where to advertise. Delaware, among states, scores high on dominance (the Philadelphia ADI nearly covers the state). In contrast, no single ADI comes close to dominating Wyoming. It is easier to allocate advertising dollars when dominance is high, it is easier to target news-making events when dominance is high, and voters receive more uniform political information when dominance is high.

The final constraint media markets place on campaigners is cost. The more costly advertising is, the more a candidate has to spend in order to inform and influence voters. I will experiment with three measures of cost. *Total cost* is the amount required to swamp a political district. Its calculation is straightforward:[1]

$$\mathrm{TC}_D = \sum_{i=1}^{n} \mathrm{Cost}_{\mathrm{ADI}_n}$$

Per capita cost is total cost divided by the district population; thus, it can be interpreted as the per citizen cost of swamping a district in advertising.[2]

---

1. Cost is the average cost of a one minute spot during the local news in each ADI. These figures were provided to my by Charles Stewart. He only has figures for the top 200 ADI's. I impute the final 13 by substituting the value from the next smallest ADI in the state.

2. My measure of "per capita cost" is mathematically equivalent to what Stewart and Reynolds call "total cost." They describe total cost as "the amount of money required to but a one-minute television spot during the local evening news to reach 1% of the TV Households in the state." (1989, 16) This conflates two things, the total cost of advertising in a state, and the cost adjusted to a per person figure. Total cost and per capita cost are not the same thing; the degree

Average cost takes into account contiguity values as well as cost. Not all advertising reaches potential voters—Las Vegas and Reno television is seen by some California, New Mexico, and Utah residents. This waste (from a political perspective) drives up the cost per Nevadan. Average cost weights per capita cost by the degree of overlap. Note that:

- Cost is the cost of a 30-second advertising spot on the six o'clock news.
- $\text{ADI}_n \cap P_D / \text{ADI}_n$ is the proportion of the ADI population that is in the district. The reciprocal is the weight applied to advertising cost. When there is a lot of waste, the proportion is low, the reciprocal is large, and the cost is inflated.
- $\text{ADI}_n \cap P_D$ is the number of individuals in the ADI—the political district intersection. Its reciprocal converts cost into a per capita figure.

Putting these three terms together results in the average cost measure:

$$\text{AC}_D = \sum_{i=1}^{n} \text{Cost}_{\text{ADI}_n} * \frac{1}{\text{ADI}_n \cap P_D / \text{ADI}_n} * \frac{1}{\text{ADI}_n \cap P_D}.$$

Data Sources

I rely on three sources for television market data. The *Television and Cable Factbook, 1989* has maps of each state with ADIs drawn in, along with TV household counts by county. I purchased an ADI County Information data set from Arbitron, Inc., which has a record for each county in the United States (except for independent cities in Virginia, which are collapsed into the surrounding counties). Each record contains the state and county Federal Information Processing System (FIPS) code, ADI (name and numerical code), TV household count, and total household count. With proper aggregation and file matching, this data set can be used to calculate state TV households, ADI TV households, and TV households in the state-ADI intersections. From the Bureau of the Census, I obtained person counts for counties and congressional districts. From the state-level tapes, I was able to construct a county-to-CD conversion scheme (more on this later). Once this was attached to the Arbitron data set, I could proceed in a similar fashion to compute CD population, ADI population, and the population of CD-ADI intersections.

---

of overlap between ADI's and districts varies dramatically. My "total cost" (as I use it) and "per capita cost" measures more clearly distinguish the two.

TABLE A.1. Arbitron TV Households and Census Person Counts

	TV Households
Census persons, county level	.994
Census households, county level	.996
Arbitron estimate of households	.999
Census persons, state level	.992
News mix (persons)	.999
Dominance (persons)	.998

Note: The first four entries are the correlation between the Arbitron TV household count and a variety of other person and household measures; the last two entries are correlations between TV market variables calculated for states using census person counts and Arbitron TV household counts.

TV Households versus Population

Stewart and Reynolds (1989) use Arbitron TV household counts to calculate contiguity, dominance, and their version of total and average costs. I rely on census population counts instead. The main reason is a practical one: TV household counts are only available at the county level, and aggregating from county level to CD is very complicated.

No information is lost if Census population estimates are used instead of Arbitron TV household estimates. Census and Arbitron estimates correlate at .99 and above. More importantly, contiguity and dominance, estimated for states using population or TV households, correlate nearly perfectly (see table A.1). The advantage of being able to compare CDs and states on identical measures far outweighs any additional errors introduced by using census population counts instead of Arbitron TV household counts.

The following are sample calculations for California's Sixteenth District (see table A.2 for data):

$$\begin{aligned}
\text{Dominance} &= \sum_{i=1}^{n}\left(\frac{\text{ADI}_n \cap P_D}{P_D}\right)^2 \\
&= \left(\frac{22,303}{525,893}\right)^2 + \left(\frac{503,590}{525,893}\right)^2 \\
&= .9188
\end{aligned}$$

$$\begin{aligned}\text{News mix} &= \sum_{i=1}^{n} \frac{\text{ADI}_n \cap P_D}{P_D} * \frac{\text{ADI}_n \cap P_D}{\text{ADI}_n} \\ &= \frac{22,303}{525,893} * \frac{22,303}{454,129} + \frac{503,590}{525,893} * \frac{503,590}{503,590} \\ &= .9597\,.\end{aligned}$$

Demographic Measures

All the demographic data were drawn from the 1970, 1980, and 1990 censuses of the population, congressional-district-level extract (state data are simply aggregated from the CD level). All demographic measures were recoded as follows (table references are to tables in the census documentation):

TABLE A.2. Sample Calculations for Media Market Measures

District/State	Pop_D	ADI Name	ADI_n	$\text{ADI}_n \cap \text{Pop}_\text{D}$
Sixteenth CA	525,893	Salinas -Monterey	503,590	503,950
		Santa Barbara	454,129	22,303
Maine	1,124,660	Portland/ Poland Spring	791,572	728,494
		Bangor	304,835	304,835
		Presque Isle	91,331	91,331
Nevada	800,493	Las Vegas	471,815	471,815
		Reno	364,585	303,242
		Salt Lake City	1,615,422	25,436
Wyoming	469,557	Casper- Riverton	124,917	124,917
		Denver, CO	2,293,099	108,964
		Salt Lake City	1,615,422	95,749
		Cheyenne/ Scottsbluff	106,993	68,649
		Rapid City, SD	216,166	37,462
		Billings/ Hardin	208,812	21,639

Source: 1980 census (population figures) and Arbitron, Inc.

- *Urban residence* proportion urban, table 1, category 2, divided by total person count (hereafter, *total person count* refers to the 100 percent count of persons, table 3, which the census uses as the base for most tables, unless specifically noted otherwise).
- *Race,* proportion black, table 12, category 2 (black), divided by total person count.
- *Foreign stock*, percentage foreign stock, table 33, category 4 (foreign born), divided by total person count.
- *Education,* percentage high school, table 50, categories 1–2 (elementary through high school, one to three years; high school, four years), divided by persons 18 years and over.
- *Occupation,* proportion blue collar, table 66, categories 6, 10–13 (private household occupations [6]; precision production, craft, and repair [10]; operator, fabricators, and laborers [11–13]), divided by employed persons 16 years and over.
- *Income,* median income, table 69.
- *Population Density,* total person count divided by square miles (coded by author).

Diversity Measures

Lieberson's (1969) A_w, often described as a measure of heterogeneity within a population, has been used by a number of political scientists (Sullivan 1973; Bond 1983; Bullock and Brady 1983). It can be used with polytomous variables.[3] The variable runs from zero to 1.0, where zero means that every individual is different and 1.0 means that there is no variation (on the characteristic under examination).[4] The advantage of this measure is that it can be used for polytomous nominal variables. Any other calculation would have required me to either make unsupportable assertions about ordinality or collapse categories into dichotomies (e.g., white/non-white).

To illustrate how these scores are calculated, suppose I am comparing religious diversity within a state. If I randomly pair all members of the state, one way to think about diversity is to see it as the proportion of pairs that differ in religious affiliation. Another way to express this is to suppose I randomly choose a pair of individuals. What is the probability that they will have different religions? For one (hypothetical) variable with three categories, this calculation is straightforward:

---

3. As noted in the text, variance is a good measure of diversity using continuous variables.

4. If, for example, I calculated diversity of social security numbers, A_w would take on a value of zero.

Catholic	=	.25	Catholic-Catholic pairs = $.25 * .25$
Protestant	=	.50	Agreeing pairs = $.25^2 + .5^2 + .25^2$
Other	=	.25	Different pairs = $1 - (.25^2 + .5^2 + .25^2)$

The diversity score, A_w, is calculated thus: given a variable, X, with n categories, and each category expressed as its proportion in the population (X_n),

$$\sum_{i=1}^{n} X_i = 1.0$$
$$A_w = 1 - \sum_{i=1}^{n} (X_i)^2 .$$

When more than one variable is under consideration, the situation is less obvious. In particular, how do I evaluate partial matches? Suppose I have three variables, with values $X = 1,2,3$, $Y = 1,2$, $Z = 1,2$. The combination (111)(111) is obviously a perfect match, but what about pairs such as (111)(112) or (111)(123)? I need to adjust for partial matches by 1) summing up the probability of exact matches, and 2) summing up all the probabilities of all partial matches, but at the same time, however, 3) weighting partial matches by the proportion of specified characteristics shared by the pair. Thus,

- C_i is the proportion of the population in some combination (e.g., 111, 121, 223, etc.)
- n is the number of possible combinations
- C_{ij} is a pair of combinations expressed as the cross-product of their proportions in the population
- W_{ij} is the proportion of specified characteristics shared in C_{ij} (e.g., for [111][112], $W_{ij} = 2/3$)

For this example, n = (the number of categories in X) $*$ (the number of categories in Y) $*$ (the number of categories in Z) = 12. Notice that $\sum_{i=1}^{n}(C_i)^2$ is the proportion of perfect matches and $2 * \sum_{i=1}^{n} \sum_{j>1}^{n} C_{ij}$ is the proportion of partial matches. A 2 is required because each match can occur twice, that is, (111)(122) and (122)(111). The final measure is calculated thusly:

$$A_w = 1 - \left[\sum_{i=1}^{n} (C_i)^2\right] + \left(2 * \sum_{i=1}^{n} \sum_{j>1}^{n} C_{ij} W_{ij}\right) .$$

This measure is: “nicely interpretable in probability terms, since it represents the proportion of characteristics upon which a randomly-selected pair of

TABLE A.3. Demographic Diversity in Election Districts: Updating Bond

Variable	States	CDs	K-S Statistic
Bond summary measure	.424 (.033)	.389 (.060)	2.469 (.001)
Summary measure 2	.590 (.031)	.586 (.038)	1.370 (.047)
N of cases	50	435	—

Source: 1980 census.
Note: The entries in columns 2 and 3 are means, with standard deviations (i.e., the variance of the variances) in parentheses. The Bond summary is a diversity measure that considers race, occupation, foreign stock, and urbanness together (replicating Bond 1983). Summary measure 2 is a summary measure that includes all the measures in this table and does not collapse across categories (as Bond does). Mathematically, including more categories results in a higher (more diverse) score.

individuals will differ, assuming sampling with replacement" (Sullivan 1973, 70). Still, the diversity measure could be deceiving. Although it represents the underlying construct—diversity of interests in a district—it is probably too catholic about the political impact of various categories. Using this measure, a district that is 90 percent white and 10 percent black receives an identical score as a one that is 90 percent black and 10 percent white. If heterogeneity and homogeneity were the *only* concern here, this result would be fine.

However, I selected the demographic categories I did precisely because I know certain cleavages are politically more potent than others. The reader should be aware of this limitation of the diversity measure. Bond (1983) and Bullock and Brady (1983) both used a summary diversity measure. For purposes of comparison, I report updated summary scores in table A.3.

APPENDIX B

Variable and Model Information for Chapter 5

Variable Names and Sources

Challenger and Incumbent Spending

- Source: Federal Election Commission tapes, released through the Interuniversity Consortium on Political and Social Research.
- Recode: Any spending below $5,000 by a candidate was set at $5,000, following Jacobson 1985. I transformed to a per capita variable by dividing by state or congressional district population. I then logged in order to account for extreme values.

Closeness

- Source: The *Congressional Quarterly Weekly Report* election reports.
- Recode: The original variable runs from zero (safe Democratic) to 7 (safe Republican). I folded the variable about its midpoint so that it ran from zero (safe seat) to 4 (too close to call).

Ideological Extremity

- Member score on the D-Nominate scale developed by Poole and Rosenthal (1985, 1997) folded about its midpoint.

Partisan Split

- Source: Election results from assorted sources.
- Recode: The variable is originally coded as a normal vote variable, or the average percentage Republican vote for the last three elections. For Senate races, I took the average value of the normal vote in all congressional districts in the state for three previous elections. This variable ran from zero to 1.0. I subtracted .5 from the variable, so that it ran from $-.5$

(perfect Democratic) to zero (even) to .5 (perfect Republican). For Republican incumbents, I left the scale as it stood. For Democratic incumbents, I multiplied the scale by −1. Now the scale runs from −.5 (perfectly favorable partisan balance for the challenger) to zero (even) to .5 (perfectly favorable partisan balance for the incumbent).

Media Contiguity, Media Dominance

- See appendix A.

Income and Education Variance

- See Appendix A.

Racial, Urban, and Foreign Stock Diversity

- See Appendix A.

Stability of Coefficients across Institutions

As mentioned in chapter 5, a "fully-dummied" specification is analogous to a Chow test for the stability of coefficients across the Senate and House. The test involves a straightforward difference in F-statistics, comparing the constrained to the unconstrained model.[1]

The results for tables 5.1, 5.2, and 5.3 are shown below. Please note that a *non-significant* p-value indicates that the unconstrained model (separate relationships for House and Senate) provides no statistically discernible improvement in fit and is thus rejected. For the final model (table 5.3), I report the F-statistic both for a "fully-dummied" specification and a specification with a single Senate dummy variable. Only the model with a separate slope provides a significantly improved fit.

Tested Model	F-statistic	(d.f. numerator/ denominator)
Table 5.1	.55	(10/2442)
Table 5.2	1.0813	(5/2452)
Table 5.3	.922	(14/ 2434)
Table 5.3 (separate constant)	3.45	(13/2435)

1. No difference in coefficients implicitly assumes that all the dummy variable coefficients are zero. The unconstrained model allows different coefficients across the Senate and House.

Problems of Two-Stage Estimation

Serious sample problems undermine any study of the causes and consequences of campaign spending in both the House and Senate. The first problem is well recognized: many cases are lost because the instrumental variables (lagged spending) are missing. This study is especially prone to this problem because I compare across the two institutions. The maximum number of Senate races in any election year is 34. In a two-stage setup, I rely on lagged values in order to solve the endogeneity problem (at least for incumbent spending). This raises a number of quandaries. First, what to do with first-term incumbents or incumbents who were unopposed last time? Gary Jacobson set up a number of screening criteria for inclusion in his campaign-spending analysis of the House (1985). The incumbent has to face opposition at time t and time $t-1$. The incumbent at time t has to be the same incumbent as at time $t-1$ (spending in an open-seat race does not reflect a "generalized propensity to spend"). This eliminates unopposed incumbents at time t and time $t-1$, all open-seat races from $t-1$, and all incumbent defeats from time $t-1$. Furthermore, a scattering of cases drop out due to retirements and deaths during the term and House members who retire or are defeated in a primary at time t. One result of this filtering rule, ironically, is that just those seats in which we are most interested—the most competitive—are eliminated from the analysis.

The consequences of the screening rule for the House are perilous. In 1978, for example, the number of races involving incumbents was 377 and the number of incumbents facing opposition was 327. Various additional reasons to drop cases reduced the working sample size to 236, resulting in a 28 percent reduction in the sample (see Jacobson 1985, 1990a; and Green and Krasno 1988, 1990). The consequences of the screening rule for the Senate are far worse. The population of Senate cases is already small. More cases are lost through filtering than for the House because Senate incumbents are defeated more often in primaries, Senate incumbents are defeated more often in the general election, and senators retire at higher rates (due perhaps to the longer term, a desire to run for governor or president, and the age of politicians who finally make it to the Senate). Whereas around 30 percent of the House cases drop out during the 1980s and 1990s , nearly 50 percent of the Senate races drop out, and this figure would have been even higher were it not for the unusually high number of cases that met Jacobson's criteria in 1988.

Incidentally, the small number of Senate cases is not a problem unique to this study. Richard Fenno, when attempting to reproduce *Home Style* for the Senate, apparently found that senators were too idiosyncratic to allow much generalization about constituency representation, campaign strategies, and Washington behavior. Instead of a Senate *Home Style*, he has written a series of books about individual senators. Canon, in his study of amateurism, could not produce some measures for the Senate because of instability due to small *n*s (Canon 1990, chap. 3). Much of the reason for the Senate being the forgotten side of Congress is numerical: there are not enough Senate races in a single election, or even a whole decade, to allow for very sophisticated quantitative analyses.

A more serious problem, however, is less frequently recognized. The cases that are excluded from the analysis are not a random sample of districts. Since they are districts held by freshmen or those where an incumbent was defeated in a primary, they are also districts where, on average, challengers would be expected to spend more and receive more votes than average. This could bias the estimates for campaign spending in unknown ways, and not just in this study. Most studies of campaign spending and incumbent vote totals encounter this potential bias.[2]

Two-stage estimation places severe demands on the data and the specification than does ordinary least squares, demands that typical congressional campaign data sets may not meet. As Bartels (1991) has shown, parameter estimates are heavily affected by specification error. I have the advantage in this case of relying on a well-established causal framework, one that is given some support in Bartels's article. The potential biases and sample size problems remain. Pooled results will be dominated by the House sample (a pooled sample, with filtering, consists of 20 times as many House as Senate contests). The two-stage estimates, then, should be viewed with some caution.

2. I am grateful to an anonymous reviewer for pointing out this problem.

APPENDIX C

Variable Information for Chapter 6

Measures Added in This Chapter

Political Attentiveness

- Source: All measures in the chapter come from the National Election Study's Senate Election Study.
- Recode: This is a combination of self-expressed interest in politics, number of days spent watching the network news, and number of days spent reading a daily newspaper. Weights were obtained via a principal components analysis of these three variables and education. No rotation was applied to the eigenvectors, and I limited the number of factors to two. Education occupies a separate dimension.

Presidential Approval

- Recode: This is a recoded version of a four-point scale, where 1=strongly approve and 4=strongly disapprove. The variable was recoded to the zero to one range, and multiplied by -1 for Democratic incumbents.

Help District?

- Recode: This item asked respondents whether they thought the incumbent would be helpful if the respondent had a problem. It is a three-point scale, running from very helpful to not very helpful. It was recoded to the zero to one range, with 1 = very helpful.

Likes/Dislikes

- Recode: Respondents provided open-ended mentions of things they liked or disliked about candidates. The NES coded up to five mentions. I totaled all mentions across individuals. This variable thus runs from zero to five. For the comparisons across various types of likes/dislikes, I used the coding scheme suggested by the NES. I standardized these measures by institution when I included them in the voting models.

Contacts

- Recode: The NES asked respondents about seven types of candidate contacts. I summed the contacts across individuals.

Vote for Incumbent

- Recode: This is a combination of reported vote for the incumbent and, for those who did not vote, expressed preference.

Isolating High-Flow Campaign Environments

In chapter 6, I compared voter information and choice models for “high”- and “low”-intensity contests. In order to reassure the reader that the threshold I chose is not arbitrary, I illustrate in this appendix the impact of other selection schemes.

I selected out particular contests to look at according to two criteria: per capita spending by the challenger was greater than one (meaning that the challenger spent more than one dollar per capita) or the ratio of incumbent to challenger spending crossed a low threshold, 1.5 and 1.0. The first criteria sets too high a barrier: too many cases are screened out, and it ignores spending by the incumbent. Campaign intensity is a *combination* of spending by both candidates along with candidate quality. [1]

As an illustration of the impact of various filters, I present the number of races, the number of respondents, and incumbent and challenger recall under three screens. These are shown in table C.1. The first thing to note is that significant numbers of respondents are experiencing a hard-fought House contest, even in a single election year.[2] In the third column, for example, 19 percent of

1. I exclude from consideration races where the incumbent spent less than 20 cents per capita, or about $100,000 for the House (this additional criterion had no impact on the Senate). This eliminate those few races where the challenger and the incumbent spent the same amount, but the campaign was a low intensity contest overall. By doing so, I increased the proportion of cases that fell into Krasno’s “high intensity” category from 52 percent to 80 percent. This lends some construct validity to the ratio measure.

2. This figure is higher than might be expected using the typical NES post-election study because the 1992 NES/SES, in order to produce satisfactory samples within each state, has more sample strata at the sub-national and sub-state level.

TABLE C.1. Isolating High-Flow Campaign Environments

	All Races	High Challenger Spending	Ratio < 1.5	Ratio < 1.0
Races				
House	317	10 (3%)	25 (8%)	8 (3%)
Senate	27	5 (18%)	7 (26%)	5 (18%)
Respondents				
House	1588	74 (5%)	305 (19%)	87 (6%)
Senate	2538	476 (19%)	433 (17%)	298 (12%)
Recall name				
House incumbent	.292	.270	.331	.356
Senate incumbent	.402	.457	.448	.460
House challenger	.108	.189	.203	.184
Senate challenger	.220	.363	.356	.409

Source: 1988 NES/SES.
Note: High challenger spending is defined as races in which challenger spending is higher than one dollar per capita. Entries for recall and recognition are the proportion correctly identifying the name of the candidate. If per capita incumbent spending is less than .2 (or 20 cents per citizen), the case is dropped. This only affected the House sample. See chapter 6.

the House respondents and 8 percent of races fell into the high-intensity category compared to 17 percent of the respondents and 26 percent of the races.[3] "Significant overlaps" in Senate and House races is more than just an empty phrase; there are hundreds of respondents in the 1988 NES/SES who experienced hard-fought House races. While sample sizes of contests become quite low, sample sizes of respondents may not.

The information measure, candidate recall, behaves as expected: recall rates increase as the filter becomes progressively finer. Challenger quality also increases; when the spending ratio is at 1.5 or less, average Senate challenger quality is 3.38 (out of 4.00) and House challenger quality is 2.59. Still, this duplicates the results of chapter 3: even in comparably intense contests with comparable levels of spending (per capita), Senate challengers are more prominent and have more political experience than their House counterparts do. Based on table C.1, as well as readings of campaign coverage and some additional analyses, I chose to label those cases in which the ratio of incumbent to challenger

3. I am fudging on the definition of "intensity" only for purposes of illustration. In previous chapters, I set this to be either a) Krasno's and Westlye's qualitative coding or b) a spending ratio less than 2.0. For purposes of illustration, a more restrictive ratio is more revealing.

TABLE C.2. Voter Choice: Defection toward and away from the Incumbent

	Defect Toward	Defect Away
Constant	−0.545 (.133)**	−1.337 (.190)**
Presidential approval	0.642 (.070)**	−0.355 (.104)**
Help district?	−0.093 (.125)	−0.562 (.185)**
Intensity	0.034 (.102)	0.219 (.146)
Institution	−0.103 (.087)	−0.007 (.135)
Incumbent likes	0.136 (.041)**	−0.327 (.093)**
Incumbent dislikes	−0.204 (.075)*	0.170 (.091)
Challenger likes	−0.233 (.071)**	0.213 (.080)*
Challenger dislikes	0.078 (.066)	−0.329 (.149)*
Incumbent contacts	−0.091 (.026)**	0.079 (.040)*
Challenger contacts	0.029 (.025)	−0.091 (.040)*
Tenure	−0.000 (.000)	0.000 (.000)
$-2\ln(L_0/L_1)$	594.66	1,420.63
N of cases	1,377	1,377
Percentage predicted	78.94	94.63

Source: 1988 NES/SES.
Note: Only partisans are included in the table. The entries in column 2 are maximum likelihood probit coefficients with standard errors in parentheses. Coefficients with one asterisk are more than two times their standard errors; coefficients with two asterisks are three or more times their standard errors.

spending was no greater that 1.5 as "high intensity." This leaves enough cases and a large enough proportion of contests that any statistical estimates should not be inefficient or biased.

Alternative Models of Voting

Zaller presents voter defection models for the House (1978) and the Senate (1988). For comparison, I present in in table C.2 pooled versions of the model in equation 6.1, using incumbent defection as the dependent variable and relying on the 1988 data in order to retain comparability with Zaller (1992, chap. 10).

There are two ways in which defection might occur. First, partisans of the opposite party might vote for the incumbent, such as a self-identified Republican voting for a Democratic incumbent. I call this "defection toward the

TABLE C.3. Incumbent Vote, No Likes and Dislikes

Constant	.934 (.137)**
Party identification	1.145 (.106)**
Presidential approval	.317 (.092)**
Help district?	.790 (.092)**
Intensity	−.720 (.071)**
Institution	.013 (.010)
Incumbent contacts	.041 (.019)*
Challenger contacts	−.121 (.018)**
Tenure	−.092 (.048)
$-2\ln(L_0/L_1)$	550.2
N of cases	2,364
Percentage predicted	78.22

Source: 1988 NES/SES.
Note: The entries are maximum likelihood probit coefficients with standard errors in parentheses.

incumbent." Twenty-nine percent of the respondents reported this kind of behavior. The second kind of defection is more unusual but still a logical possibility: incumbent partisans vote for the challenger. Nine percent of the sample reported that they voted for or preferred the opposing party's challenger.

The results reinforce in most ways the conclusions in this chapter. Presidential approval, perceptions of incumbent helpfulness, and the balance of likes and dislikes are all important predictors of the likelihood of defection.[4] The results are mirror images: variables that are positively related to defection toward the incumbent (such as incumbent likes) are negatively related to defection away.

The discrepancies between these results and the results reported in the text are instructive. There is no discernible difference between the likelihood of defection in high- and low-intensity races, although the direction of the effect is positive, as would be expected, and nearly statistically significant in the "defect away" model. Voters who defect toward incumbents are basically voters who have positive things to say about the incumbent and have not had many contacts with challengers. They are, as Zaller notes, mainly individuals exposed to one-way information flows (1992). Campaign intensity has little to do with the likelihood of defection. Similarly, those who defect away mainly do so because they dislike incumbents and like the challenger, regardless of the

4. Partisanship is not in this model since partisanship crossed with the incumbent's party makes up the dependent variable.

intensity of the race. Finally, I include candidate likes and dislikes as independent variables in my models predicting the vote. I examined whether this was a misspecification: placing the dependent variable on both sides of the equation. The number of likes and dislikes that a respondent cites correlates with vote choice at .30. There is quite a bit of unexplained variation left in the vote. Second, as is shown in table C.3, there is only a minimal change in the other coefficients in the model when likes and dislikes are removed. The goodness of fit declines substantially.