5

Unraveling a Paradox

In a review of research based on the 1978 NES/CPS survey, Timothy Cook observes that "voters in general may point to personal qualifications of incumbents, remember being contacted, recall 'something special' done for the district, or expect helpful responses from the staff. Members of Congress in general may spend time communicating a personal image to constituents, intervening with bureaucracies, and protecting district interests. But it remains unclear how closely connected individual incumbents' activities are to their constituents' evaluations, and decisions."[1] Since Cook's review additional studies have appeared, but their conclusions are consistent with his. There appears to be little or no evidence that the activities and resource allocations of congressional offices have any effect on the perceptions, evaluations, and voting decisions of their constituents. Consider the following roll call of recent findings:

On campaign spending
The more incumbents spend, the worse they do (Jacobson).[2]

On travel
Incumbent electoral success is unrelated to visits home (Fenno).[3]

On staff
Incumbent electoral success is unrelated to district staff presence (Fenno).[4]

On casework

"Using returns from the 1978 Congressional elections, we estimate the effects of three strategies that incumbents may use to increase their vote totals . . . we find that casework has no statistically significant effect and further find that a substantively important effect is unlikely, given our estimates" (Johannes and McAdams).[5]

On federal spending

"All in all, we found *no* evidence that obtaining local federal spending for his district or protecting against spending cutbacks is a useful way for a congressman to pursue reelection" (Feldman and Jondrow).[6]

On legislative activity

Incumbent electoral success is unrelated to bill sponsorship and committee position (Ragsdale and Cook).[7]

On campaign communications

"Voters' knowledge of candidates' personality characteristics is not increased (even among those voters who are most likely to pay attention to political communications) when candidates emphasize their personality in their campaign communications . . . Indeed, the findings presented here suggest that when politicians talk, nobody is listening" (Raymond).[8]

On resources in general

"Of paramount interest, incumbent resources, including representative activities and campaign money, do not significantly affect the district vote" (Ragsdale and Cook).[9]

"As for the ultimate question—how to explain the incumbency effect and the diminishing marginals—we cannot claim to know the final answer. Casework, it seems, is not part of the answer. Neither, apparently, are many elements of the incumbent's 'home style'—trips home, allocating staff resources to the home district, and the like" (Johannes and McAdams).[10]

Does incumbent behavior affect voter perceptions and behavior in more subtle ways, such as by attracting campaign contributions or discouraging strong challenges? Apparently not:

"Generally, the evidence does not support the hypothesis that bringing government spending or employment to his district is a good way for an incumbent to raise campaign funds or to avoid a serious challenge."[11]

"Incumbents' use of the 'perks' available for 'advertising and casework,' on the other hand, is not related to any indicator of

> challenger quality . . . the use of these 'perks' does not discourage campaign contributors nor politically experienced opponents once the effects of ideological discrepancy and partisan vulnerability have been taken into account [perks include district offices, total staff, percentage of staff in district, and percentage of staff designated as caseworkers]."[12]

When viewed in combination, these findings are troubling.[13] An intelligent layman perusing the political science literature could arrive at the remarkable conclusion that House incumbents can close their district offices, fire their staffs, stop doing casework, abandon the quest for federal money, give up their district residences, choose committees purely on the basis of personal interest, tell the political action committees where to put their money, and still get as many votes—perhaps more—than if they continued to behave as incumbents presently do. No doubt politicians in general and incumbent congressmen (and their challengers) in particular would regard this conclusion as preposterous. Indeed, we are sure that with the possible exception of Johannes and McAdams none of the researchers cited would accept the stark implication of their combined findings, though they accept their individual findings considered in isolation. Only Fenno and Jacobson hesitate to take their findings at face value.

When research results conflict with theory, common sense, or even political folklore, scholars should entertain the possibility that something could be wrong with the research.[14] In the present context, rather than blithely concluding that high-level politicians have no contact with reality, scholars should pause and wonder whether something might be wrong with the data and/or methods. In the course of our research we have encountered scores of nonrelationships and even a few perverse negative relationships between incumbent efforts and constituent perceptions or evaluations. This repeated confounding of expectations has at least three bases. The first simply involves the crude and imprecise nature of the data. The second involves problems of simultaneous causation which render suspect ordinary statistical procedures. The third reason seems obvious in retrospect but has been overlooked heretofore. It involves problems of cross-level inference. We illustrate the arguments with congressional data because the prevailing doubts arise from analysis of such data, but the arguments apply equally to the British case and indeed to any studies that attempt to determine the response of the citizenry to the behavior of political leaders.

Problems in Relating Member Behavior to Constituent Reports

Problems of conceptualization and measurement pervade attempts to link incumbent behavior to constituent perceptions and evaluations, and lead directly to problems of specification in statistical analyses. Some variables are easy to measure—the number of district offices and campaign expenditures, for example. Others present greater difficulties. Anyone who has asked congressional staff how many cases their offices handle in a month, how many grant applications they monitor, what percentage of time the member devotes to constituent service, and so forth is well aware of the uncertainty attending responses to such questions. Estimates are data and better than nothing, but it is dangerous to treat them with the confidence we might treat something like dollars of expenditure. Attempts to relate the quantitative information supplied by staff to reports by constituents often come to naught. For example, the bivariate relationship between number of cases handled per week by a congressional office and constituent evaluations of the congressman's constituency attentiveness is essentially zero. When "caseload" is included in a multivariate equation, the coefficient is often tiny and statistically insignificant. Yet recoding the variable in a manner that places less weight on small differences in caseload levels produces a different finding. Table 5.1 provides a vehicle for some extended discussion. The third column of the table differs from the second only in the inclusion of some individual level variables, a matter not addressed until later. For now, simply compare column one with two.

The dependent variable is the individual response to the survey item, "Do you happen to remember anything special that your U.S. representative (name) has done for this district or for the people in this district while he/she has been in Congress?" Some of the district level variables relate to the individual responses in the expected way. For instance, constituent recollections are significantly higher in districts whose offices report an active effort to get grants, whose congressmen come home frequently (weekly or more often), and whose congressmen are Democrats. Constituent recollections are significantly lower the more junior the congressman. This is a reasonable finding if constituents in fact remember achievements from longer ago than the very recent past. Caseloads, however, do not

Table 5.1. Correlates of district service recollections (probit estimates)[a]

Variable	(1)	(2)	(3)
Frequent visits	.318**	.344**	.371**
Democratic incumbent	.217**	.211*	.281**
No. of district offices	−.068†	−.020	.003
Office seeks grants	.258**	.225**	.188**
Year elected	−.025**	−.022**	−.018**
Cases per week	.000	—	—
Medium caseload	—	−.043	−.050
Heavy caseload	—	.143†	.204*
No. of staff members	.022	—	—
Medium staff	—	−.128	.133
Large staff	—	.353*	.273†
Impersonal contact	—	—	.966**
Second-hand contact	—	—	.724**
Constant	.369	.356	−1.040*
Chi-square/df	62/7**	75/9**	271/11**

a. $n = 1591$. $^{**}p < .01$.
$^{*}p < .05$. $\dagger p < .10$.

seem to matter in Equation 1. The coefficient suggests not the faintest trace of a relationship. But consider Equation 2, where caseload is specified not as a continuous variable but by dummy variables representing low (up to 30), medium (31–90) and high (over 90) levels. Low and medium levels are indistinguishable, whereas high caseload levels relate significantly to high recollections of district service. Here the substantive conclusion hinges on a specification decision.

Social scientists are loath to throw away information in the data, but the contrast between the first and second columns of the table may indicate that the data contain less information than meets the eye. Possibly the caseload estimates are better used to separate legislators into broad groupings rather than to distinguish small differences among them. With estimated caseloads so uncertain, the definition of a case not constant from office to office, and the mix of cases (such as individual versus high level) unmeasured, should we really expect one office's figure of 50 to be associated with a constituency response significantly different from another office's figure of 60? Probably not. But such an expectation might be more reasonable when comparing figures of 30 and 100.

On the other hand, the coefficients of the staff variable suggest additional possibilities.[15] Administrative assistants would be expected to provide a more accurate report of the full-time equivalents in their office than of their weekly caseload, but the coefficient of the continuous measure of staff (Equation 1) is small and insignificant, whereas a discrete classification (Equation 2) produces a stronger relationship, as again, only offices with larger than normal staffs can be differentiated from the other categories. The raw data show a clumping of offices in the range of 8–11 staff members; of 87 offices whose congressmen sought re-election, only 13 had fewer than 8 staff members, and only 8 had more than 11. There is no discernible relationship in the large middle category, but there is a discernible one on the extremes, namely, a doubling in various reports and evaluations as one moves across the three categories (Table 5.2). Large staffs are associated with the effect for district service recollections and expectations of helpfulness, whereas small staffs are associated with the effect for positive comments about the member's constituency attentiveness. Multivariate analyses analogous to those in Table 5.1 confirm these bivariate impressions.

The estimates for staff possibly indicate the inappropriateness of a linear specification. Alternatively, the results may reflect no more than noisy data, as with caseloads. At first glance, measuring staff is relatively easy: count the bodies. But staffs are more or less efficient, and more or less geared to district service. Considerations like these may muddy the import of the precise figures, whereas broad categorizations may separate offices into meaningful categories.

The general point is not that categorical representations of incumbent resources and activities are superior in any ultimate sense

Table 5.2. Size of Washington staff and constituent responses

	No. of staff members		
Constituent response	0–7	8–11	12+
Remembers district service	16%	20%	31%
Incumbent would be very helpful if problem arose	22	27	42
Positive mention of incumbent attentiveness to constituency	8	14	16

to the continuous measures.[16] Rather the point is that caution is needed. If conclusions are contingent on the precise mode of representing the variable of interest, then no analysis can be considered conclusive until there is a consensus on the most appropriate measure and specification.

Many researchers act as if such a consensus were at hand. For example, almost every relevant analysis thus far published or presented has assumed that the effects of incumbent resources and activities are linear. Thus, the typical regression analysis requires that the estimated effect of 20 trips home is $20b_i$, of 35 trips $35b_i$, of 50 trips $50b_i$, and so forth. Such linearity assumptions are untenable in a wide variety of circumstances. Consider representative A in the protectionist career stage gliding along with election victories of 70 percent.[17] The possible electoral benefit of any particular activity is limited by the existing electoral base since anyone who would ever consider voting for the representative already may be a supporter. In contrast, consider representative B, a young member in the expansionist stage trying to make a marginal seat safe. This representative might get considerable mileage out of behavior which would only keep representative A in a steady state. The positions of some members leave room for their activities and allocations to have larger effects than those allowed by the positions of other members. The use of logit and probit methods goes some way toward incorporating such considerations because these methods presume diminishing effects of the independent variables as the probability of the dependent behavior approaches unity.

Congressmen in the same electoral circumstances may have constituencies that differ in other relevant respects. For example, representative C's caseload consists largely of individual requests from socially marginal people, such as the elderly poor, disabled veterans, and immigrants, who vote at low levels, give little money, and remain outside of the established social networks. In contrast, representative D has a smaller caseload, but it contains more requests from businesses and organized groups whose members vote, contribute, and communicate with other political actors. The prosperous constituents of representative E ignore the announcement of $500,000 in federal funds, whereas the less fortunate constituents of representative F rejoice at the announcement of a grant of $300,000. Representative G, a busy subcommittee chair, cannot get home as often as representative H, a freshman with few pressing Washington

responsibilities, but G compensates by means unavailable to H and unmeasured by us.

In short, the effects of incumbent activities and resource allocations are naturally conditional upon various features of the representative's district and political circumstances. It would be convenient if every variable of political interest worked in simple, additive, unconditional fashion, but it is unlikely that they do. The problem for researchers is that representatives themselves are much more likely to understand the true forms of the relationships and the elements that condition those relationships. Choices that are perfectly comprehensible to them pose a complicated problem for analysis. Because this problem does not yield to the simplest attempted solution, however, does not mean that its structure is purely random.

Problems of Simultaneous Causation

The problems discussed thus far can be addressed by renewed efforts at data gathering, hypothesis formulation, and statistical analysis. While labor intensive, such problems are less intractable than problems of simultaneous causation. To illustrate, consider Table 5.1 once again. Note that the district offices variable takes on a negative sign twice, and in one case achieves statistical significance. As the number of district offices goes up, recollections of district service, positive comments about incumbents, expectations of helpfulness, and overall job ratings go down. In multivariate estimations *significant* negative relationships occur with some regularity. Thus, to Jacobson's strange observation that "the more incumbents spend, the worse they do," we can add the equally odd observation that the more district offices incumbents have, the less their constituents like them.

Understandably surprised by his finding, Jacobson went on to explore its basis. The results are now well known. Incumbent spending is principally defensive; it is highly responsive to challenger spending. Incumbents do worse when they spend more precisely because spending more usually indicates that they face a strong challenge which makes them do worse.[18] This circularity lies at the root of the statistical problem. Let us examine the district offices anomaly more closely. Table 5.3 presents data on the most striking example: as the number of district offices goes up, expectations of helpfulness decline in close to a negative linear fashion.

Table 5.3. Confusing causal linkages: an illustration using district offices

No. of district offices	Very helpful	Don't know how helpful	No. of respondents
1	35%	18%	774
2	27	26	716
3	20	27	428
4	15	32	104
5	9	30	23

Year elected	Very helpful	Don't know how helpful	Avg. no. of district offices	Avg. 1976 vote
1964	29%	20%	1.6	62.4%
1964–1972	29	21	1.9	64.2
1974	26	25	2.2	58.8
1976	22	35	2.4	59.9

A larger number of offices is strongly associated with lower expectations of helpfulness. In addition, the more district offices representatives have, the higher the proportion of their constituents who do not know how helpful they would be. The bottom panel of the table contains some additional information. Less senior members have more district offices; they also have a less secure electoral base than more senior members and a less well-formed reputation in their constituencies. Thus, a second hypothesis emerges. New members of Congress who are less well known, less positively evaluated, and less electorally secure establish more district offices in order to improve their political standing. Their efforts may well have the expected positive effect, but that effect may be overwhelmed in cross-sectional data by the initial weakness (relatively speaking) of newer members.[19]

If a relationship is composed of two partially offsetting components, one would like to break down the relationship into the separate components. In essence, two-stage estimation procedures using instrumental variables attempt to do just that. By substituting an instrument for the problem variable on the right-hand side, one attempts to obtain an estimate of the one-way effect of the right-hand side variable on the left-hand side variable. In practice, diffi-

culties often arise because of the requirement that the variables comprising the instrument be related to the right-hand side variable but not the left-hand side variable. Previous electoral margin, for example, relates significantly to number of district offices, but it also relates to most of the variables one would use district offices to predict, so it is not appropriate to include margin in an instrument. Thus, in attempting to "purge" a right-hand side variable of its troublesome variation, one may end up purging it of all its systematic variation. After reading earlier versions of Johannes and McAdams and Feldman and Jondrow, one of us raised in each case the likelihood of simultaneous causation.[20] Johannes and McAdams responded with a two-stage analysis that supported their simpler analyses. Unfortunately, the instrument for caseload explains about 5 percent of the variance in the original caseload variable, so it is essentially a different variable, or more likely, sheer noise. We have constructed an instrument for district offices using the regression:

$$\begin{aligned} Y = {} & .053 \text{ year elected} - .012 \text{ city} - .011 \text{ suburb} \\ & - .456 \text{ Democrat} + .456 \text{ subcommittee chair} \\ & - .146 \text{ party leadership} - .138 \text{ college} - .805, \qquad R^2 = .29 \end{aligned}$$

When substituted in the equations, the coefficient on the instrument is almost always insignificant ($-.043$, s.e. $= .089$ in Equation 1 of Table 5.1), but rarely positive, and never positive and significant. This may indicate either that the number of district offices truly has no positive impact on constituents' perceptions and evaluations or that the instrument poorly represents the true effects of district offices.

In early stages of research, analysts may hesitate to get involved with full-blown simultaneous equations structural systems and the time, effort, and money they entail. But they should realize that their analyses are "first cuts" which do not provide findings so definitive as to call into question the behavior of politicians who have achieved the highest levels of elective office. In an uncertain world some aspects of political behavior may well reflect superstitution more than reality. But the faith put in simple statistical methods does not appear to be altogether divorced from superstition.

Overlooked Difficulties in Cross-Level Analysis

Imagine a hypothetical society in which there is a perfect one-to-one relationship between the efforts of representatives and the re-

sponses of their constituents. To be exact, if a representative devotes x percent of his or her time to constituency service, x percent of his or her constituents respond with a favorable rating or expectation. Thus, by definition, the aggregate regression for the population shows a constant of 0, a slope of 1.0, and an R^2 of 1.0.

Suppose that political scientists now take a survey of this society. If they are both very careful and very lucky and draw perfect samples within a legislative district sampling frame, each district sample will mirror the district. To provide some numbers for analysis, assume that the total sample consists of 9 district samples having 10 constituents each (Table 5.4).

If the district average rating is regressed on percentage of legislator time, the estimates remain as before—constant of 0, slope of 1.0 (both with standard errors of 0), and an R^2 of 1.0. Generally, analysts do not aggregate survey data, however, preferring to make use of the larger number of individual observations. In the example, if the 90 individual observations are entered rather than the 10 aggregated district observations, the resulting regression is:

$$\text{Percentage of approval} = .00 + \underset{(.18)}{1.00}\ (\text{percentage of time}), \qquad R^2 = .27$$

The regression parameters have not changed, but the switch to individual level analysis has resulted in a drastically lowered estimate of explained variance, and a positive standard error for the coefficient of percentage of time. The reason is that the regression equation makes only 9 distinct predictions, one per district: .1, .2, .3, .4, .5,

Table 5.4. Hypothetical data set

	Constituent approval		
District	Yes (= 1)	No (= 0)	Legislator time
1	1	9	10%
2	2	8	20
3	3	7	30
4	4	6	40
5	5	5	50
6	6	4	60
7	7	3	70
8	8	2	80
9	9	1	90

.6, .7, .8, .9. Thus, *everyone* in a given district is predicted to have the same value. Moreover, since all constituents actually have values of either 0 or 1, *no one* is correctly predicted by the estimated value for those in their district. The important feature of this example is that 27 percent is the *maximum* variance that any district level variable can explain even though, by construction, the district level relationship is perfect.

Recall the "fundamental identity" of analysis of variance: total variation equals between-group variation plus within-group variation. In the hypothetical example only 27 percent of the variation is between-district, and the district level variable explains all of it. Fortunately, there is still a large, significant coefficient on percentage of time, which indicates the importance of the variable, but conditions need not be so ideal as the example so far presumes. Suppose that political scientists were not quite so lucky and picked up a 10 percent error in their approval ratings through the vagaries of the data collection process.[21] Using a table of random numbers, the approval rating of one constituent of the ten in each district is reversed. The regression now becomes:

$$\text{Percentage of approval} = .18 + .70\ (\text{percentage of time}), \qquad R^2 = .12$$
$$(.19)$$

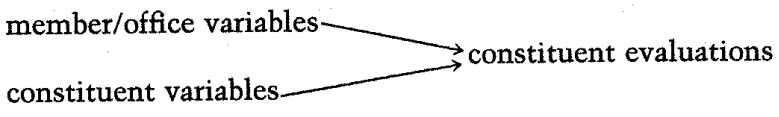
and the estimated importance of percent time has dropped 30 percent. Measurement error in the right-hand side variables of a regression equation biases the coefficient estimates, with the bias taking the form of underestimation in the one variable case.[22] Thus, studies of the effects of member activities and resource allocations using rough estimates of caseloads and crude proxies like number of staff generally will misrepresent the effects of those activities and allocations. And while important, that is not the only point of the example.

Refer back to Table 5.1. The estimated equation reported in column three differs from those in columns 1 and 2 by the addition of two powerful individual level right-hand side variables. As shown, constituents who report contact with their congressman through newspapers or magazines, TV, radio, or mail are significantly more likely to remember something special the congressman has done, as are those who have heard about some friend's or relative's contact. Notice the veritable explosion in the chi-square statistic between columns 2 and 3. The individual level variables obviously make a huge contribution to the explanatory power of the equation. This

should come as no surprise. The 1500 people in this analysis fall in only 70 districts; on average, 23 people *are assigned the same value for any given district level variable.* Any variation among that score of constituents is completely beyond the explanatory reach of the district level variable. Given the relative similarity of resource usage among congressmen, such as staff size, within-district variation should normally far exceed between-district variation. Since restricting the variance of the right-hand side variables inflates the estimates of their standard errors, analysis of member activities and allocations using crude measures, taken from a limited sample of districts, encounters a kind of statistical "double whammy" that biases coefficient estimates while simultaneously inflating their standard errors.[23]

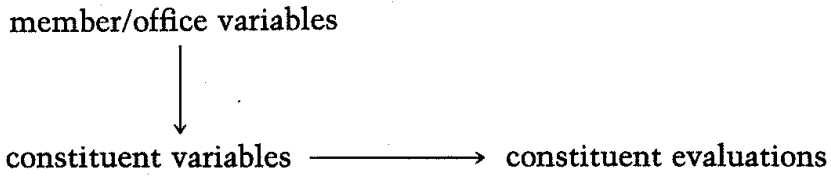
But if within-district variation dwarfs between-district variation, isn't that an important finding, and doesn't it show that what members do and decide actually does have only minimal consequences for their reputations in the district? For a number of reasons this negative implication is unwarranted. In the first place, the analysis of variance analogy breaks down, since this is not an experimental situation. In an ideal world the congressional offices in the sampling frame would be divided into two groups, and one group would close down all its district offices, reduce its staff to a secretary and receptionist, and quit seeking grants, while the second group would expand its offices, staff, and activities to the legal maximum.[24] This experimental design would legitimately test for the existence and magnitude of treatment effects. But as matters actually stand, few incumbents are either very much lower or very much higher than the rest, as shown by the size of the staffs in Washington. Thus, even if member activities have a large potential for affecting their constituents' perceptions, the observable effects are not likely to be large if members do not differ greatly in what they do.

There is still another reason why the relative magnitudes of within-district and between-district variance can be misleading. The comparison would be important if the following model were the correct one:



Here, the member and constituent variables are theoretically and statistically independent, and given comparable measurement quality, if the latter explain the lion's share of the variance in constituent

evaluations, one might reasonably conclude that the former are of little significance. But how would member activities and decisions have effects except *through* constituent variables? How would high caseloads, big staffs, grant seeking, and frequent visits produce positive evaluations if constituents did not read or hear about the congressman, see him or her on TV, get something in the mail, or recognize the name? Thus, the model most researchers implicitly have in mind is



In this case it would not be surprising if member activities made no contribution to explanations of constituent evaluations, since the hypothesized effect is indirect, *through* other constituent variables. From this standpoint the estimations in column 3 of Table 5.1 are somewhat puzzling. Why do frequent visits, high caseloads, large staffs, grant seeking, and Democratic incumbency remain significant when contact variables are added to the analysis? Though the wording of the latter appears comprehensive, they apparently do not fully capture the ways in which incumbent activities affect constituent evaluations.

The point of this chapter is not to attack the usefulness of statistical methods. Rather, the point is to identify a number of reasons why simple applications of common methods may go astray. If a simple tabular presentation or parsimonious regression strikingly bears out a hypothesis, that may be sufficient, though analysis should not necessarily stop there. But if simple methods do not bear out what logic, common sense, previous findings, traditional wisdom, or plausible theories predict, analysis should certainly not stop there. Although data themselves are neutral, every data analysis implicitly makes a host of assumptions about the structure of the data and the relationships among them. If those assumptions are misconceived, the conclusions based on the data may be correspondingly misconceived.