SOCIAL PSYCHOLOGY, DEMOGRAPHIC VARIABLES, AND LINEAR REGRESSION: Breaking the Iron Triangle in Voting Research

Christopher H. Achen

A previous paper showed that a simple prospective model of voting and party identification subsumed much of the social-psychological and retrospective voting literatures, in the sense that it rigorously implied their key findings and added many new ones as well. This paper extends the argument by showing that the same prospective voting model has drastic implications for conventional statistical specifications in voting research. First, linear models should be discarded in favor of a particular nonlinear specification. Second, demographics should be dropped from the list of independent variables.

In political science, quantitative researchers enjoyed their first and most enduring success in voting studies. After the invention of survey research, empirical generalizations poured forth: in no other field do we have so many hard facts with which to discipline our thinking. In most American presidential years, for example, we can forecast the aggregate national vote with astonishing accuracy several months in advance (e.g., Rosenstone, 1983). We know which issue positions are best predicted by income, which by education, and which by race or religion, and we know how these correlations have changed over time. There is a rapidly growing body of comparative international data as well; properly measured, many of the concepts developed in America fit other nations as well (Johnston, 1988). Factual disagreement within the field is very modest by social science standards: Good data make for consensus.

Statistical methods have undergirded the achievements of opinion research. Indeed, linear regression and its extensions have so dominated the journals that they became identified with "good work." For many years,

Christopher H. Achen, Department of Political Science and Institute for Social Research, University of Michigan, Ann Arbor, Michigan 48106.

voting research had an almost formulaic character: In a national survey sample, the vote was expressed as a simple additive funciton of party identification (PID), issues, and demographics. Regression might be replaced by probit or two-stage least squares, but the underlying linear framework of party ID, issues, and demographics was almost never questioned.

The statistical practices of voting researchers were driven by the social-psychological thinking that initially captured the field. Group identification was said to structure individuals' thinking, and the stronger the identification, the stronger the effect. Essentially, being the child of Democrats led to being a Democrat oneself, and being a Democrat led to liberalism (Campbell, Converse, Miller, and Stokes, 1980, chapters 6 and 7). Exogenous changes in group membership, such as geographic relocation or divorce, might alter party ID, but opinion changes were not likely to do so (Campbell et al., 1980, pp. 146–150). Hence it was sensible to treat party ID as exogenous, and to "explain" the vote as a function of it, along with demographics and issues as controls for group loyalties and short-term forces.

Such a framework is clear enough, perhaps, but it is less than adequate in the face of changing political realities. Group loyalties are largely stable, but in actual elections the effects of party ID, group identifications, and contemporary issues wax and wane. How are these changes to be accounted for? When and why does party identification change over time? How does it interact with issues? When might one or the other dominate the vote? How precisely do voters compromise among many different group influences? How do these decisions vary with the political life histories of the voter and the nation? For all too many of these questions, social psychology had no answers.

The central difficulty is that the causal connections subtly intertwine. If we understood them, powerful explanatory devices would lie in our hands. At minimum, some rough sense of their interaction is critical to carrying out dependable empirical work. But the interconnections were left unspecified by the social-psychological framework bequeathed to voting research. In that tradition, every causal impact was investigated separately, as if the systemic whole did not exist. Not surprisingly, in spite of sustained effort by many researchers and a considerable body of discrete experimental discoveries, little theoretical integration occurred. (Important exceptions include Jackson, 1975; Franklin and Jackson, 1983; and Brady and Sniderman, 1985.)

In this intellectual climate, nearly all researchers defaulted to the simplest possible statistical specification: "ordinary" linear regression (i.e., linearity with an assumed constant causal effect for each independent variable). Thus, the statistical models assumed what everyone knew to be false, namely that each causal effect operated independently of the values of the

other variables. Researchers studying the vote knew that campaign stimuli influenced Independents more strongly than party identifiers, so that campaign effects are not linear, but they used linear regression models with fixed effects anyway. Researchers studying party ID knew that experienced voters had more stable PIDs, so that lagged PID effects are not linear, but they used linear regression models with fixed effects anyway. Researchers studying vote defection from party ID knew that the best-informed and the worst-informed voters were more faithful to their party ID than the middling-informed, so that information effects are not linear, but they used linear regression models with fixed effects anyway.

The benefits of linear empiricism are not negligible, especially when compared to no empiricism at all, but they are soon exhausted. In voting studies, the point of diminishing returns arrived perhaps a decade ago. After the many discoveries and successes of the sixties and seventies, no one any longer expects that the next election will bring new findings about anything but itself. Too many regressions have already been run.

This disappointing sequence of events recurs across the discipline (and outside it as well). As Charles Ragin (1989) notes, quantitative social scientists are obsessed with ordinary linear regression equations, in which the marginal effect of every independent variable is forever fixed and their joint impact is simplistically additive. The actual social world, he says, is quite different. There effects are "contextual" (nonlinear in each variable) and "holistic" (interactive): The marginal impact of a variable depends both on its own value and on that of other causal factors. Forcing an oversimplified linear model on the data benefits us little, Ragin says, and in field after field, the Procrustean methods of much "large N" research account for their wattery findings and feeble theoretical impacts.

Ragin's arguments are very familiar, indeed nearly cliches. Moreover, statistical methods for interactive nonlinear specifications have long been known. Yet Ragin is surely right when he argues that, just as in the voting field, in practice nearly all social scientific statistical specifications are garden-variety linear-in-variable models.[1] One might add that Ragin's point is just as true of probit and logit, selection bias models, time series applications, their multiequation extensions, and other developments of the last two decades: The higher-powered methods and more intensive computing add little when applied to unsophisticated linear-in-variable specifications.

For all its factual contributions, linear empiricism bears a certain share of the blame for the modest contribution of voting research to contemporary political theorizing. The numbing repetition of the same statistical specifications has discouraged theoretical speculation and driven lively minds to more topical matters. Thus, every election brings forth a new round of studies: Why did Candidate X win? Which issues were influential? How

much difference did the economy make? Did the campaign matter? How was the party ID balance affected? Such questions are not uninteresting, of course, and their great advantage is that they are descriptive features of a single election and therefore can be studied within the confines of a national probability cross-section. Their great disadvantage is that they are of limited theoretical interest and must be answered all over again as soon as the next election rolls around. The journalistic nature of the enterprise escapes no one for long, and malaise soon sets in.

Like many before it, this paper proposes the obvious alternative, that theory be taken seriously. That bromide turns out to have real meaning in this context. Indeed, it has radical consequences for the day-to-day working habits of opinion researchers.

As an illustration, this paper specifies a very simple but powerful model of voter behavior, namely rational voting with incomplete information. Then statistical specifications for explaining party ID and the vote are derived from the model.

Drastic conclusions follow. First, Ragin is right: Linear regression and probit models are misspecified and should be ignored. In particular, voter information is theoretically critical, and without particular nonlinear controls for it, no statistical estimates of vote or party ID equations should be trusted.

Second, when researchers are being theoretically serious, demographics should be discarded. They belong neither in party ID nor in vote equations. The voter's political history is the only causal variable. Age, social class, and other background factors will be correlated with history, of course; they may provide a serviceable summary for purely descriptive purposes. But they do not belong in explanatory equations.

The next section sets out the intuitions behind the argument. The following sections outline the model, list the conclusions that follow from it, and explain how econometric work may be done within its structure.

SUBSTANTIVE ASSUMPTIONS

The model in this paper rests on three principle substantive assumptions:

1. *The voters are rational*. This postulate means that voters do not ignore the information they have, do not fabricate information they do not have, and do not choose what they do not want. Thus, the voters need be neither geniuses nor saints. They are required only to do their best with the information they have. It suffices, for example, that they take the advice of better-informed friends or groups.

Obvious as it sounds, the rationality assumption contrasts sharply with other perspectives on voting. For example, the voters in the social-psychological tradition tend to follow the party identification of their parents, whether it makes sense for them or not. And they may imagine that the candidate of their party is closer to them than is actually the case, simply to make themselves feel better by reducing cognitive dissonance. In the rational choice tradition, such behavior is explained as the inevitable sampling errors stemming from reasonable inferential rules.

2. *The voters are prospective in orientation.* This postulate also contrasts with other familiar viewpoints, including some occasionally mistaken for rational choice models. "Retrospective voters," for example, choose at least partly on the basis of the past, even though they cannot hope to change it and its lessons may have long ceased to be relevant. For the rational voter, however, bygones are bygones, and only the future matters. Thus, retrospective voting is not a rational choice theory. As standard decision theory teaches, the rational chooser looks forward, not backward. The rational voter is a prospective voter, and the past is useful only for its clues about the future. (Recent evidence for prosepctive voting includes Lewis-Beck, 1988, and Lockerbie, 1989.)

3. *There are just two parties and they each offer benefits to voters that oscillate over time around a fixed but unknown mean.* Unlike the first two assumptions, which would obtain in any model of this kind, this third one specializes the model to a simple case for ease of analysis—a stable two-party system. For each voter, the parties offer benefits, sometimes rather more, sometimes rather less. The benefits vary independently from one term of office to the next around a central tendency. Thus, the parties may oscillate left or right by chance, but they do not drift steadily in any one direction. In effect, the model describes a stable period between realignments.

Within the overall framework, one might analyze voter behavior under quite different assumptions, for example, that the parties trend over time, undergo realignments, and so on. The arguments would be quite similar. But for present purposes, the fixed mean and independent sampling assumptions are maintained.

In this setup, only the true mean benefit is of interest to the voter. She wishes to forecast future benefits so that she can vote intelligently, and party benefits are drawn independently over time from a common distribution. Hence to predict benefits to come, she can do no better than to forecast the mean. Since that mean is unknown, however, she must use her past experience with the parties to estimate it. Under the simplified model of this paper, all past benefits are a random sample from a common distri-

bution. Hence from elementary statistics, the stylized rational voter knows that a simple average of past benefits is the best estimate. It is convenient to think of her as computing this average as each term of office ends.

For voters with a good deal of experience, that ends the story. A new voter, however, has no such past experience, and a young voter may have very little. Such voters will not wish to place too much emphasis on the current campaign or the most recent president, since their information is so noisy. Hence they will bring a "prior distribution" to bear: beliefs about the parties that predate the voter's initial entry into the political system. The most likely source for these beliefs is their parents.

If parents and children tend to occupy similar social positions, and if party benefits are related to social positions, then it is rational to use parental party ID to help form one's own party ID. One will not hold the ID as strongly as the parents, because one's own life will not be identical with theirs. But the voter makes fewer mistakes, especially in the early years, by using parental ID as a starting point. Over time, as will be seen below, parental influence wanes as political life experience accumulates.

This model, then, falls into the voting research tradition begun by Key (1966), who argued that elections constitute evaluations of incumbent performance, especially of the economy (see also Downs, 1957). For Key, party identification was a rational choice, a "standing decision" subject to modification with further experience (Jackson, 1975; Fiorina, 1981). Rational-choice theorizing along these lines includes models of individual voters by Zechman (1979), who studied single-election campaign dynamics, and by Calvert and MacKuen (1985), who treat dynamics across realignments. By contrast, this paper and its predecessor study voter dynamics *across* elections and *within* realignments. This focus leads more easily to macro-level results, such as the dynamics of the party distribution over time. It also matches up more simply with the available data, which consist primarily of single-election cross-sections within a party alignment system. In consequence, many more analytic results are possible than with prior approaches, most of them verifiable with existing data.

THE MODEL

We now set out the model more precisely. First, note that in a two-party system only the difference in benefits between the parties is relevant to the voter's decision. As noted above, the model specifies that voter i's party difference in benefits at any given time t, denoted by u_{it}, varies randomly around its mean:

$$u_{it} = \delta_i + v_{it} \tag{1}$$

where δ_i is the voter's mean or long-term benefit differential and v_{it} is the deviation from the mean at a particular time t. It is assmed that v_{it} has mean zero and finite variance $\omega_i^2 > 0$, and that it is distributed independently over time.

We also assume that across the population the true distribution of party benefits is normal. Any distribution could be used as a test case; the normal is analytically convenient. Since the scale of party benefits is arbitrary (up to an affine transformation), we take the distribution to be standard normal.[2] Thus, writing $f(\delta_i)$ for the density over the population, and $N(\mu, \sigma^2)$ for the density of a normal distribution with mean μ and variance σ^2:

$$f(\delta_i) = N(0, 1) \tag{2}$$

At time 0, the voter enters the political system. She has no direct knowledge of δ_i, her benefit difference between the parties. However, she has a prior belief about it, derived from parental experience. It was shown in Achen (1989) how such a distribution may be derived. In effect, children compromise between their parents' PID and the population average—a form of regression to the mean. The more social mobility in the society, the less relevant parental experience is likely to be, and hence the more the children will discount it. Similarly, erratic benefit streams during the parents' lifetimes also lead to their children discounting parental experience. In all cases, the children are assumed to behave as sensible Bayesians, and the voter's initial PID turns out to be a linear function of her parents' PID.

As succeeding periods pass, the voter gets experience with the parties. She receives one observation per period from equation (1), each independently distributed.[3] She uses this information to optimally update her beliefs, converging toward the "right" PID (cf. Franklin, 1984, p. 474).

The relevant Bayesian theorem is the following:

Theorem. Suppose the scalar random variable x is distributed $N(\mu, \sigma^2)$ and that n independent realizations $x_1, x_2, \ldots, x_n$ are observed. Denote the sample mean of the x_i by $\bar{x}_n$. If μ has a prior distribution $N(\mu_0, \sigma_0^2)$, and if σ^2 is known, then the posterior distribution is $N(\mu_n, \sigma_n^2)$, where:

$$\mu_n = \frac{h_0\mu_0 + h_l\bar{x}_n}{h_0 + h_l} \tag{3}$$

with "precisions" $h_0 = 1/\sigma_0^2$ and $h_1 = n/\sigma^2$, and where, setting $h_n = 1/\sigma_n^2$:

$$h_n = h_0 + h_l \tag{4}$$

This theorem applies to the voter in the following way. At time 0 voter i has a prior belief about her party benefit differential, derived from her

parents, normally distributed with mean $\hat{\delta}_{i0}$ and variance σ_{i0}^2. After n periods, she also has experience with the parties (her likelihood), summarized by the mean value of her benefit streams, $\bar{u}_{in}$ with variance ω_i^2/n. Applying the theorem produces the voter's current benefit estimate, namely $\hat{\delta}_{in}$, along with its posterior variance:

$$\hat{\delta}_{in} = \frac{h_{i0}\hat{\delta}_{i0} + h_{il}\bar{u}_{in}}{h_{i0} + h_{il}} \tag{5}$$

where as above $h_{i0} = 1/\sigma_{i0}^2$, $h_{il} = n/(\omega_i^2)$, $\bar{u}_{in} = \Sigma_{t=1}^{n} u_{it}/n$ and the precision of current PID is $h_{in} = h_{i0} + h_{il}$.

This equation (5) defines the "current party identification" of voter i: It is her current estimate of the benefit differential between the parties, namely $\hat{\delta}_{in}$. This quantity is an estimate of the underlying true value, δ_i, which is called the "long-term party identification." These definitions suggest how one might integrate the older notion that a voter has a stable PID that generates a "normal vote" with the newer perspective that she also has a PID that she continually revises as the political life of her country evolves.

Equation (5) has a "retrospective" form: The voter estimates current PID from a weighted average of parental PID (captured in her initial PID, $\hat{\delta}_{i0}$) and her political experience (summarized by $\bar{u}_{in}$).[4] But the logic is rigorously prospective: Both parents and personal history are useful just to the extent that they predict the future, and not otherwise. In fact, the retrospective appearance is an accident of the special case considered here, in which party benefit streams are constant over time. If the model were applied to the period of a realignment or to parties whose platforms and appeals are changing, it would cease to be retrospective.

The elements of equation (5) vary across individuals and ordinarily cannot be measured directly. For example, the voter may well be uncertain about the mean party benefit she has experienced over her lifetime. Thus, in practice, it must be expressed as a function of other measures, and the result should be substituted into equation (5). Typically, several such substitutions will be necessary so that unknown parameters will appear in both numerator and denominator. In addition, all the measures in use will be noisy variables derived from surveys. The resulting nonlinear expression will behave during estimation like a sensitive and overwrought child under disciplinary interrogation. Not much good sense will ensue, and the wise will avoid the task entirely.[5] A better approach is to use prior PID to incorporate parental effects, and then to express current PID as a function of partisanship at the prior period plus the most recent retrospection. From equation (5), we have that the voter's PID at the previous period was:

$$\hat{\delta}_{i,n-1} = \frac{h_{i0}\hat{\delta}_{i0} + h_{il'}u_{i,n-1}}{h_{i0} + h_{il'}} \tag{6}$$

where $h_{i0} = 1/\sigma_{i0}^2$ as before, $h_{il'} = (n - 1)/(\omega_i^2)$, $\bar{u}_{i,n-1} = \Sigma_{t=1}^{n-1} u_{it}/(n - 1)$ and $h_{i,n-1} = h_{i0} + h_{il'}$.

Substitution into equation (5) yields:

$$\hat{\delta}_{in} = \frac{h_{i,n-1}\hat{\delta}_{i,n-1} + h_{\omega i}u_{in}}{h_{i,n-1} + h_{\omega i}} \tag{7}$$

where $h_{\omega i} = 1/\omega_i^2$ is the inverse of the sampling variance of the nth observation.

This equation expresses party ID as a function of retrospective benefits since the last period plus PID at that period. It is far more manageable than equation (5), since its elements are more easily measured. Even with good measures, however, nonlinear estimation of Bayesian models has proved difficult in practice; few examples are in print. As Bartels (1991) notes, these models are nearly unidentified. In equation (3), if σ_0^2 and σ^2 are approximately proportional across the population, meaning that voters' initial variance in their benefit estimates is proportional to the variance in perceived benefits during their life (which would occur under the model if parents tend to have approximately the same perceived variation in benefits as their children), then the weights are also proportional. Hence numerator and denominator have a common factor of σ^2, and the latter parameter is not identified.

Under the proportionality assumption, multiplying numerator and denominator in equation (7) by ω_i^2 cures the unidentifiability. Entering parameters for scale differences among variables and assuming that each variable's zero point corresponds to Independence, we then have an approximate PID equation of the form:

$$PID = \frac{\beta_1(RE) + (\beta_2 + \beta_3 n)(PID_{-1})}{\alpha + n} \tag{8}$$

where PID_{-1} is the lagged value of PID, RE is the retrospective evaluation for the past term of office (which might consist of several variables, each with its own coefficient), and n is the number of terms experienced by the voter.[6] This is a mildly nonlinear equation, easily estimable with contemporary software. It is the basic statistical specification for party identification that follows from the model of this paper.[7]

Finally, it is assumed that in the course of a campaign, the voter gets

some additional information $c_{i,n+1}$ beyond her prior party ID about what $u_{i,n+1}$ is likely to be in the next period:

$$c_{i,n+1} = u_{i,n+1} + \epsilon_{i,n+1} \tag{9}$$

where $\epsilon_{i,n+1}$ is normally distributed with mean zero and variance τ_i^2. The term $1/\tau_i^2$ will be denoted by $h_{i\tau}$.

Combining current PID plus campaign information, voter i estimates benefits in the next period $(u_{i,n+1})$ as:

$$\hat{u}_{i,n+1} = \frac{h_{ic}\hat{\delta}_{in} + h_{i\tau}c_{n+1}}{h_{ic} + h_{i\tau}} \tag{10}$$

where h_{ic} is the inverse of the sum of the posterior variance of current PID plus ω_i^2.[8]

Hence the voter chooses the first party if its expected benefits exceed those of the second party, that is, if:

$$\hat{u}_{i,n+1} > 0 \tag{11}$$

This equation may be estimated by nonlinear probit analysis under the assumption that omitted variables are normally distributed with mean zero and the same variance for the population. The right-hand side of the probit equation is that of equation (8), with an additional term in the numerator for current issues. Adding intercepts to capture the means of unmeasured effects yields:

$$\alpha_0 + \frac{\alpha_1 + \beta_1(RE) + \beta_2(I) + (\beta_3 + \beta_4 n)(PID_{-1})}{\alpha_3 + n} \tag{12}$$

where all variables are defined as before and I is the set of variables measuring current campaign issues.

It is important to note that both equation (8) for PID and equation (12) for the vote are functions of party identification *at the prior period*. The current party identification recorded in the midst of the campaign incorporates retrospective evaluations; it is a consequence of this model that, with sufficiently good measurement, retrospections should drop out when *current* PID is used to predict the vote.

VALIDATING THE MODEL

The model has now been set out. Like any other new perspective, its value is best assessed by establishing its fit with what we already know.

Eleven propositions follow rigorously and straightforwardly from its as-
sumptions (Achen, 1989). Here we simply list them, along with references
to the empirical literature where the effect has been verified:

Proposition 1. New voters' current PIDs are initially positively corre-
lated with that of their parents but are more centrist (Jennings and Niemi,
1981, pp. 90–91, 153).

Proposition 2. New voters' current PIDs are initially more labile (higher
subjective variance of estimate) than their parents (Jennings and Niemi,
1981, pp. 50–51).

Proposition 3. New voters with more labile parents have more labile
PIDs; they will also have more centrist PIDs than other new voters whose
parents have the same mean experience with the parties (Jennings and
Niemi, 1981, pp. 50–51, 86–87).

Proposition 4. The effect of parental PID on children's current PIDs is
positive, but it declines over time at a diminishing rate (Jennings and
Niemi, 1981, p. 90; this result was also derived by Calvert and MacKuen,
1985).

Proposition 5. Relative to the impact of parental PID, the effect of total
life experience on voters' current PIDs is positive and increases with time.

Proposition 6. The effect of lagged PID on current PID increases with
time; relative to lagged PID, the effect of current benefits diminishes with
time (Markus, 1979; Franklin and Jackson, 1983).[9]

Thus, the young moved disproportionately toward the party of successful
incumbents in the 1930s (Campbell, Converse, Miller, and Stokes, 1980
[1960], pp. 153–157) and the 1980s. Moreover, the adjustment in party
identification occurs at the end of the first term. Thus, major changes to-
ward the Democrats should occur in 1936 rather than in 1932 (Campbell,
Converse, Miller, and Stokes, 1980 [1960], pp. 534–535) and toward the
Republicans in 1984 but not in 1980 or 1982.

Proposition 7. (Key's standing decision model). Party identification
changes only when benefits in the current period are unexpectedly high or
low relative to the last period's current PID; the effect of the innovations
diminishes over time (MacKuen, Erikson, and Stimson, 1989).

Proposition 8. If all voters have the same posterior variance in initial PID
and in benefit estimates, if the intraperiod voter benefit levels are not neg-
atively correlated, and if the mean voter is a nonpartisan, then partisanship
grows with time (Converse, 1969; Campbell, Converse, Miller, and Stokes,
1980 [1960]:162–163).[10]

Proposition 9. Assume the conditions of Proposition 8, but let informa-
tion about party benefits (w_i) vary across the population independently of
mean benefits (δ_i). Then on average partisans are better informed about
their party benefits than nonpartisans (Campbell, Converse, Miller, and
Stokes, 1980 [1960], pp. 142–145).

Proposition 10. If two voters have the same posterior variance for current PID and the same variance in campaign information, then the more Independent of the two will respond more strongly to campaign information (Converse, 1966b).

Finally, one may also derive the well-known curvilinear relation between campaign information and defection from party ID:

Proposition 11. Controlled for current PID, its posterior variance, and the over-time variance of party benefit differential, the probability of defection is lower at high and low levels of information and is highest at intermediate levels (Converse, 1966c; see also Geddes and Zaller, 1989; Zaller, 1989).

In summary, then, the simple model proposed here subsumes within a unified framework a broad range of findings in the literature. When the model is used to structure statistical work, there is no need to worry each time about whether a specification for changes in party identification, let us say, is consistent with what we know about low-information voters, Independents, and retrospective voters. The model automatically applies the findings of prior work and avoids theoretical inconsistencies. The price, however, is a radical redirection of conventional work habits, as the next section demonstrates.

ECONOMETRIC CONSEQUENCES FOR PARTY ID EQUATIONS

Theoretical clarity has dramatic repercussions in empirical voting research. While many of these have been noted above, and the rest are implicit in the equations derived, it may be useful to spell out the consequences individually. This section begins with the equations that explain party identification.

1. *Demographics do not appear in theoretical equations explaining party ID or the vote.* This proposition states what is obvious from inspection of equations (8) and (12). PID and votes are explained by beliefs about the future, and those beliefs are formed from parental socialization, political experience with the parties, and current campaign issues. Occupying a particular social niche may help us guess the voter's experiences and beliefs, but they are not themselves explanatory (Franklin, 1984, p. 467).

In practice, of course, measuring beliefs is difficult. Errors of measurement abound, and complex econometric methods often must be employed to correct for them. If demographics predict beliefs but have no causal impact themselves, then they meet certain of the criteria to be "instrumental variables," variables that can be used to "purge" noisy measures of their errors. Thus, demographics can be useful and may sometimes even be essential. But for theoretical purposes, they are just a tool, and their coeffi-

cients in auxiliary purging regressions are of no interest. They do not belong in the main explanatory equation. Only if no measures of the theoretically relevant quantities are available should demographics be used as proxies for them, and then only with the understanding that the results will be crude and ungeneralizable.

2. *There are no legitimate ordinary linear regression equations that explain party ID across a national sample: Their coefficients are biased.* In party ID equations, ordinary linear regression models apply only to groups of voters with the same level of political information and experience (Franklin and Jackson, 1983).

This conclusion applies both to equation (5), in which party ID is expressed as a function of parental PID plus the voter's experience with the parties, and also to the more statistically relevant equation (7), where PID is a function of one-period lagged PID plus the benefits of the most recent period. A glance at the latter equation will show why the proposition holds. Party ID is expressed as a ratio. So long as the denominator varies over observations, the regression is not linear. But the denominator is a function of two kinds of information: socialization from parents and political experience. Only if both kinds of knowledge are equal across the population under study will the denominator be fixed. And unless it is fixed, it cannot be absorbed into the coefficients in the numerator to make the expression linear.

The nonlinearity of the party identification equation in general samples is not merely a technicality safely ignored in practice, nor is it a flukish consequence of one arbitrary model. Instead, it expresses a fundamental and well-known fact about human beings in general and voters in particular that no theoretically serious model of PID will escape: The experienced are different from the inexperienced. While only recent history is vivid to the young, the crises of a generation ago remain present to the old, and they see current events in longer perspective. The use of ordinary linear regression to explain party identification denies this difference, and it is biased for precisely that reason.[11]

3. *There is no theoretically meaningful equation of any form that relates party ID to parental PID plus current issues.* This proposition simply restates equation (5): A voter's current PID begins with her parents and then is affected by her entire life's experience, not just her current situation. Again, the point is not technical but substantive. No sensible model will want to specify that a lower-middle-class voter with Democratic parents and moderate views who was out of work during the Depression, or fought in World War II, or attended previously all-white schools under court order, or lost an arm in Vietnam will respond to contemporary politics in just the same way as a similarly situated voter who lacks those experiences. Yet

the researcher assumes just that when ordinary linear regression is used to predict party ID from parental PID and current issue positions. Such regressions are hopelessly biased.[12] Only if the history of the voter's experience with the parties is controlled do meaningful (nonlinear) specifications emerge.

4. *Controlled for PID at the prior period plus retrospective evaluations, the effect of parental PID on current PID is exactly zero.* This statement repeats equation (7). If lagged PID and current benefit levels (issue positions) are already in an equation for PID, parental PID does not belong there. The intuition is transparent and largely independent of this model's particulars. Parental PID starts the voter out as a Democrat or a Republican. But then the voter is on her own. If she begins as mildly Democratic because her parents were, she does not grow continually ever more Democratic from her twenties on into old age due to having Democrats for forebears. Parental influence is embodied in party ID, not continually added to it.

To assume the contrary, as so many empirical studies do, is to imagine that initial parental socialization is supplemented throughout life by additional powerful impacts like those of childhood. Typically, the equation is applied to the entire population, including those 60 years of age and older, whose parents are often deceased. How, then, are the postulated socialization effects to continue? The implicit parent-child relationship seems to be Norman and his mother at the Bates Motel.[13]

5. *Party identification recorded during a political campaign incorporates both retrospective and current issue components; its coefficient is biased upward.* The "current PID" from election surveys is usually recorded immediately prior to an election. This PID is not the $\hat{\delta}_{in}$ referred to in the propositions above, but something else. It has already been noted that the use of party identification taken from the middle of the campaign incorporates retrospective evaluations. However, campaign information may also be used to temporarily update PID. Thus, there should be movement in party ID during a campaign if the value of information is unexpected (Brody and Rothenberg, 1988). The result is an overestimate of the effect of party identification.

Odd as it may sound, a better estimate of the effect of current party identification might be had by substituting a lagged value of PID, perhaps one recorded during an earlier quiet local election, as in the Butler-Stokes British election study (1969). As one would expect from theoretical considerations, the latter measure shows none of the spurious contamination by the vote that afflicts the measure taken in the midst of the 1964 campaign. Combined with retrospective evaluations and current issue positions, and estimated with due regard for nonlinearity, the result would be a more

honest estimate of the impact of party identification in the sense political
scientists usually have in mind, that is, an estimate shorn of effects due to
incumbent evaluation and campaign issues.

CONCLUSION

This paper has argued that voting research has outgrown the linear-in-
variables regression model. A simple formal model of how voters learn im-
plies a quite different statistical formulation. Moreover, in this new frame-
work, demographic variables do not appear. They may be useful in various
econometric procedures as instrumental variables—to be used and then
ignored—but they have no proper theoretical interpretation. For purposes
of genuine understanding, they are irrelevant.

These conclusions were derived from one particular model of voter be-
havior, and in one sense, they are specific to it. But there is no cheap
escape from the general point: Rigorous theorizing is needed. We need to
know not that education matters but rather what it is about education that
makes the voter choose differently. Demographics are clues, not hypoth-
eses. No model of the voter's decision process from any theoretical per-
spective is likely to bring us back to linear regression with demographics.

Carrying out the implied program for public opinion research will not be
easy. First, no one theory commands assent, and even if one were avail-
able, we are all trained to do something else. For most readers, perhaps,
all the dreary rigor will seem unnecessary.

The new program will also seem unattractive because it will make the
work harder, particularly empirical analysis. When theoretical concepts are
no longer treated as equivalent to our measurements, all statistical pro-
cedures become more sensitive. Contrary to widespread belief, random
measurement error in multiple regression can either raise or lower coeffi-
cients. There is no general tendency for error to attenuate coefficients. In
certain circumstances, otherwise-random errors correlated among them-
selves can produce statistically significant and highly inflated coefficients of
the wrong sign (Achen, 1983, 1985). Worse yet, relatively little economet-
ric analysis has been carried out for measurement error of any sort in non-
linear equations. Methodological spadework is needed.

The promise, however, is great. At present, voting research is awash in
disconnected empirical generalizations. Progress is very nearly halted. Too
often, all coefficients seem to be significant because our models are bad and
our measures are noisy: Every demographic variable we can measure has a
little pseudoexplanatory power.

In such circumstances, even crude models are invaluable. They force
attention to clearly specified causal processes. They eliminate variables,

pick a specification, instruct the researcher in estimation, and in general, constrict the latitudinarian lifestyle that has debilitated voting research thus far. Accepting that regimen is the first step toward vigorous intellectual health for voting studies.

NOTES

1. Technically, "linear in variables" means that each independent variable enters an additive regression equation once, untransformed and with its own regression coefficient—"ordinary" linear regression. In practice, of course, variables are sometimes logged or otherwise transformed to create mildly nonlinear effects. However, Ragin's argument is that this is insufficient. Hence the substantively interesting interactions and nonmonotonic effects that are commonplace in verbal social science literature rarely appear in "large N" work.
2. Note that this condition implicitly assumes that, on average, the parties are equally attractive to the population. Typically, they are not: In that case, rescale δ_i by subtracting the population average.
3. The error term in equation (1) is white noise; generalizations are quite possible but are less transparent than this version.
4. The weights are h_{i0} and h_{it}, which correspond to socialization and information, respectively. The first measures strength (not direction) of parental socialization, that is, how much information was conveyed by parents about party differentials. The second expresses the variability experienced by the voter in party benefits over time: the sum of true variability and her errors of estimate (w_i^2, which equals n/h_{it}).
5. The implication is that credible estimates of parental effects on PID are elusive. Certainly they cannot be summarized by a single number: for example, the percent of American parents who transmit their PID to their children. No such number exists. For if both parent and child have fluid PIDs over time, what could it mean to say that the child has inherited a PID from the parent?
6. Intercepts may be added in practice to deal with omitted variables. The form of the equation is that of (12) below.
7. For n, one might substitute a linear function of it, such as number of years in the political system or, if everyone enters at the same age, then age itself as in Franklin and Jackson (1983).
8. The extra variance derives from the attempt to forecast the benefits of the next period rather than their mean.
9. Franklin and Jackson find diminishing effects of current benefits only for those elections since the National Election Studies began asking voters about parties' issue positions. For prior surveys, in which voter perceptions must be inferred from other questions, they find insignificant effects (of the wrong sign).
10. For definitions of "intraperiod benefit levels" and "partisans," along with further discussion, see Achen (1989). Note that the causal mechanism is different than in Converse: The act of voting itself has no effect here, which seems to accord with empirical findings (Franklin and Jackson, 1983).

11. It should be obvious that the disease is not cured by adding variables for age or experience to the regression equation.
12. The same remark applies to equations in which the *change* from parental PID is regressed on current issue positions.
13. Of course, when noisy survey measures are used, parental party ID will have a significant coefficient when PID is regressed on it, even with lagged PID, retrospective evaluations, and issues controlled. But that is no refutation of the Bayesian model. The specification errors combine with the measurement errors to give weight to every variable, particularly since voters interpret questions about their PID in different ways (Flanigan, Rahn, and Zingale, 1989). In these circumstances, parental PID will proxy for many errors of omission and measurement, and the results will be useless. A serious test of the Bayesian model (e.g., credibly assessing whether a coefficient in a vote or party ID equation is truly zero) demands correction for measurement error. That correction will be difficult.