The Rolling Cross-Section and Causal Attribution

Henry E. Brady and Richard Johnston

For capturing campaign effects, the main alternative to the panel design is controlled daily release of sample, the "rolling cross-section." Unlike the panel, the rolling cross-section cannot by itself capture individual change, but it is a more practical and cost-effective alternative for capturing aggregate shifts. It can, moreover, be combined with a panel, such that each design adds power and precision to the other. By some substantive criteria, the rolling cross-section dominates all alternatives. All respondents are "new to the survey" so conditioning effects are minimized. The potential fine "granularity" of sample release facilitates causal attribution by making it possible to link campaign events directly with subsequent opinion change. After fieldwork is completed, at the analysis stage, the design is supremely flexible because the sample can be cut apart at any time point, whereas panels require ex ante decisions about the choice of interview dates. As a result, the design can muster far more statistical power than might first appear from the inevitably small sample collected for any given day. The unifying fact behind these advantages is that the probability a respondent will be interviewed on any given day is as much a product of random selection as is that respondent's initial presence in the sample. But the smallness of daily samples is a serious issue. The granularity made possible by continuous, unbroken, but low-intensity fieldwork comes at a price: the limited statistical power to distinguish individual days.

This essay addresses these issues. It opens with a paradigmatic illustration of causal attribution from the closely fought 2000 presidential campaign. The illustration shows the limitation of a panel design in causal attribution, but it also shows the limits of the rolling cross-section. This forces us to ask just what the rolling cross-section is and how it might be deployed, the topic of the second section. Then follows an exposition of the logic of the primary method of compensating for the potential lack of statistical power: graphical smoothing. Part of the argument is for graphs as such: the rolling cross-section makes their use both desirable and relatively unproblematic. They are desirable in that they greatly facilitate primary research—not to mention exposition—where a major element in analysis is real time. They are relatively unproblematic because of the random assignment of each respondent to an interview date; controls for respondents' accessibility are just not required. But the smallness of daily samples forces graphical data to be smoothed, and choices among smoothing alternatives are not simple. We end with the discussion of a mixed design and a quick overview of other literature about analyzing the rolling cross-section design.

An Example

Johnston, Hagen, and Jamieson 2004 argue that a pivotal feature of the 2000 campaign was a shift in perceptions of Al Gore's character, in particular of his honesty. It would be natural for a researcher to assume beforehand that one of the major campaign events causing opinion shifts would be the presidential debates and to design a panel to capture possible shifts. Figure 1 certainly points in this direction. Figure 1a sets up data from the 2000 National Annenberg Elections Study (NAES)[1] as if resources had been committed to a simple three-wave panel with interviews before the debates (September and the first two days of October), between the first and last debate (October 3 to 16), and after the last debate (October 17 to the end). Mean values for Gore's honesty rating are indicated by solid horizontal bars, with 95 percent confidence intervals around them, for each of the three periods. For interpretive ease, ratings have been rescaled to the -1 to $+1$ interval, with values below zero conveying negative judgment.[2] The narrow confidence intervals reflect the massive accumulation of sample in the NAES.

Unquestionably, Al Gore was better regarded before the first debate than after it. The predebate mean is positive while the postdebate mean is negative. The confidence intervals suggest that there is no possibility that

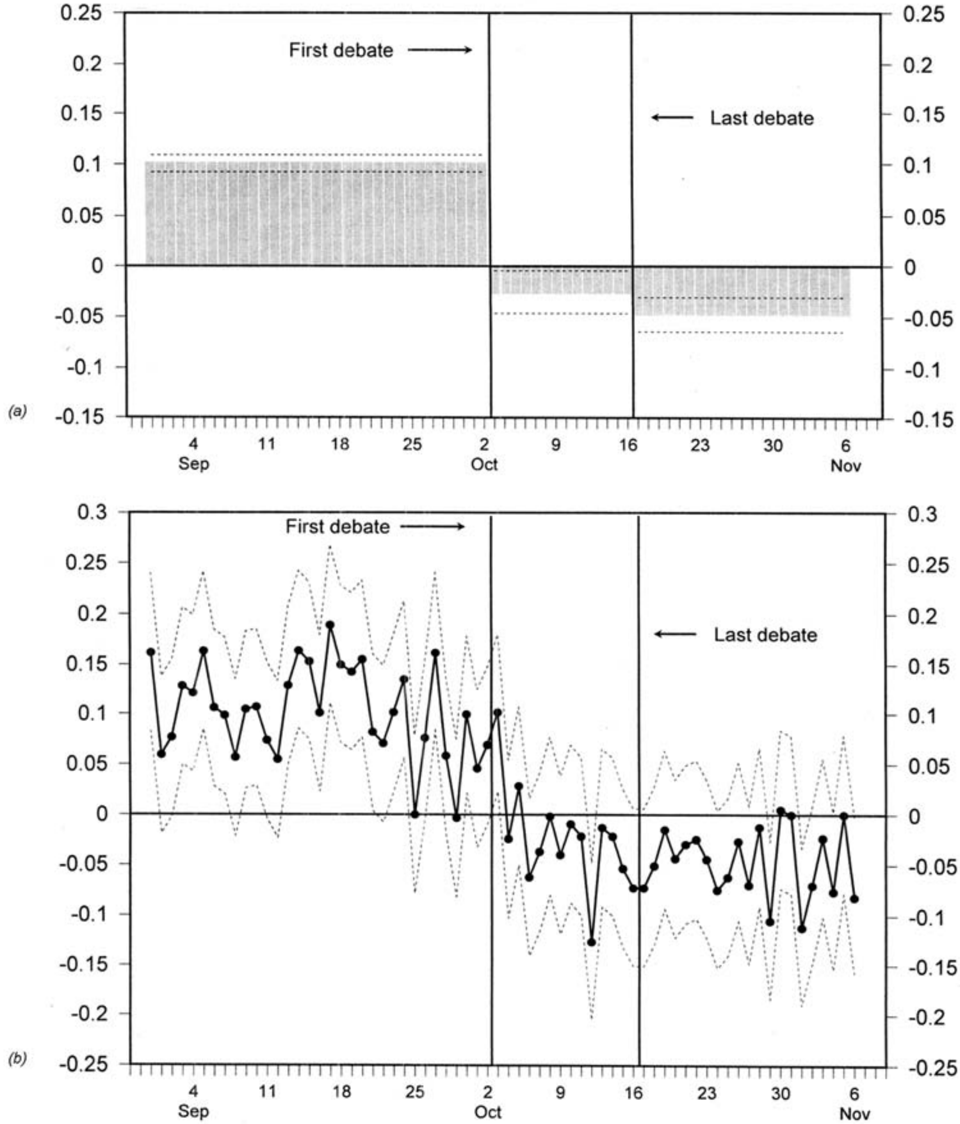

Fig. 1. Debates and perceptions of Al Gore's honesty. (*a*) Pre-post means; whole-period estimates; dashed lines are approximate 95 percent confidence interval. (*b*) Daily means; daily estimates; dashed lines are approximate 95 percent confidence interval. (Data from 2000 Annenberg Election Survey.)

these results were generated from the same underlying distribution. If any debate mattered, it must have been the first one, as the first shift is both larger and statistically less ambiguous than the second one. There is a suggestion that opinion on Gore deteriorated further after the second debate, but even the large sample sizes in the period do not allow us to reject the null hypothesis of no difference between the days before and the days after the last debate.

But how do we know that any debate was critical? The data were cut arbitrarily at the dates of the public events to simulate the results from a panel designed on the premise that debates are crucial moments in the history of a campaign. It is, of course, a reasonable supposition that if anything has dynamic impact, debates will. But the analysis is based on that supposition, not on any consideration of actual dynamics. Certainly, if one is precommitted to a panel design, it would not make sense to mark the boundary between interview and reinterview at anything other than a major public event. The campaign is about more than public events, however.

Figure 1b suggests that using only a crude pre-/postevent design might lead to an inappropriate causal attribution. In this panel the NAES data are fully rolled out as a daily tracking. The data are noisy, of course, as indicated both by the amount of surplus day-to-day vertical movement in the data and by the 95 percent confidence interval, which is an order of magnitude larger than those in figure 1a. Notwithstanding the noise, the first debate does not seem to be the whole of it. Values in the week or so before the first debate are lower than those that typify early September. There is a strong hint, then, that downward movement started even before the debates. But there is also a suggestion that emphasis on the first debate is not entirely misplaced. The day after the first debate witnessed a sharp drop in Gore's rating. The drop was not the largest of the series, but it is one of the few that were not corrected by immediately following observations. Could it be that the debate accelerated the decay? The picture also suggests that other debates did *not* affect Gore's ratings. The entire drop after the first debate occurred in the first few post–first debate days.

The picture so far is unclear. The temporally crude but statistically powerful periodization in panel a confirms that Gore's ratings dropped. There is no question that a sharp contrast exists between the period before and after the first debate. But figure 1b indicates that focus on the first debate does not do justice to the data. Movement probably predated the debate and may not have been affected by it. Then again, it might have, and identification of the real predebate turning point is next to impossible. As the rolling cross-section data are presented in panel b, they are powerful enough to undermine an exclusive emphasis on the debate but not powerful enough to underpin a conclusive alternative interpretation.

The Design

What is the design that gets us to this point? In essence, a "rolling" cross-section is just a cross-section of respondents, but with a twist. In any survey, when the list of potential respondents is released to interviewers to begin the process of contacting them for an interview, the interviewers are asked to follow a careful mix of calling at different times of the day and on different days of the week in order to maximize the chance of eventually finding the respondent at home. The process of completing interviews in this way is called "clearing the sample." Aggressive and systematic clearance compensates for the accidents of daily life that cause people to be away from their telephone at different times. Much of the variation in the quality of surveys and of survey houses lies in the willingness to spend money on clearance. As a result, any self-respecting survey will have several days for clearance built into it.

At the same time, the more such days, the more vulnerable the survey will be to changes in responses because of real events. People called by pollsters after September 11, 2001, for example, had much different attitudes on terrorism and defense than people called just before the tragic events of that day. But to complicate things, some of the apparent effect of time will not be from events in real time but from differences in the respondents: from any sampling frame, respondents interviewed later in the clearance period are likely to differ systematically from those easier to reach and thus interviewed earlier (Dunkelberg and Day 1973; Hawkins 1975; Groves 1989). Disentangling impact from factors evolving in real time from impact due to mere accessibility of respondents is a formidable task. But failure to take the task on may lead an analyst to misrepresent the data.

The rolling cross-section design converts the "bug" of temporal heterogeneity into a "feature." The steps in executing the design for a telephone survey are as follows.

1. Generate enough random four-digit numbers (married to known live exchanges and area codes, obviously) to achieve a target number of completed interviews over a specified period.
2. Divide the total body of telephone numbers into "replicates," each large enough to generate a given target of completions, where the target is the minimum number of completions for a subperiod, say, a day. The division of the total into replicates is essentially a random subsampling process.
3. Release replicates to the interviewers in a controlled fashion. This can be an equal number per day, or the number of replicates released on any given day can reflect priors on the importance of events in a period or on the "frequency domain" of political time.
4. Whatever the schedule for release, treat each replicate the same as every other. This means holding the numbers open for a specified number of days and applying a callback schedule that yields a constant clearance profile over the days that follow release to field. The callback schedule may vary over days of the week and over weeks of the year, reflecting known facts about the general accessibility of persons and households. Samples have to be worked harder on weekends and in holiday seasons, for example, to ensure equal probabilities of contact between normal weekdays and weekends or holidays.[3]

Figure 2 illustrates this for a representative day in the 2000 NAES. On July 5, 2000, enough numbers were released to complete 300 interviews, the target that was used for every day for the rest of the campaign. This represents six NAES replicates, where each replicate had a completion target of 50. The number of NAES replicates released per day reflected the intensity of campaigning: quite high during the presidential primaries, low in the fallow period of late spring, and higher than ever from July 5 to Election Day. Figure 2 shows the time path by which the 275 ultimate completions from the July 5 replicates were accomplished. Of all such completions, just over 40 percent (115 to 120 interviews) were recorded on July 5 itself. Over half of these interviews stemmed from the first call, and most of the rest took place on the second call. One number received four calls. Just under 20 percent of all completions (about 50 interviews) came the next

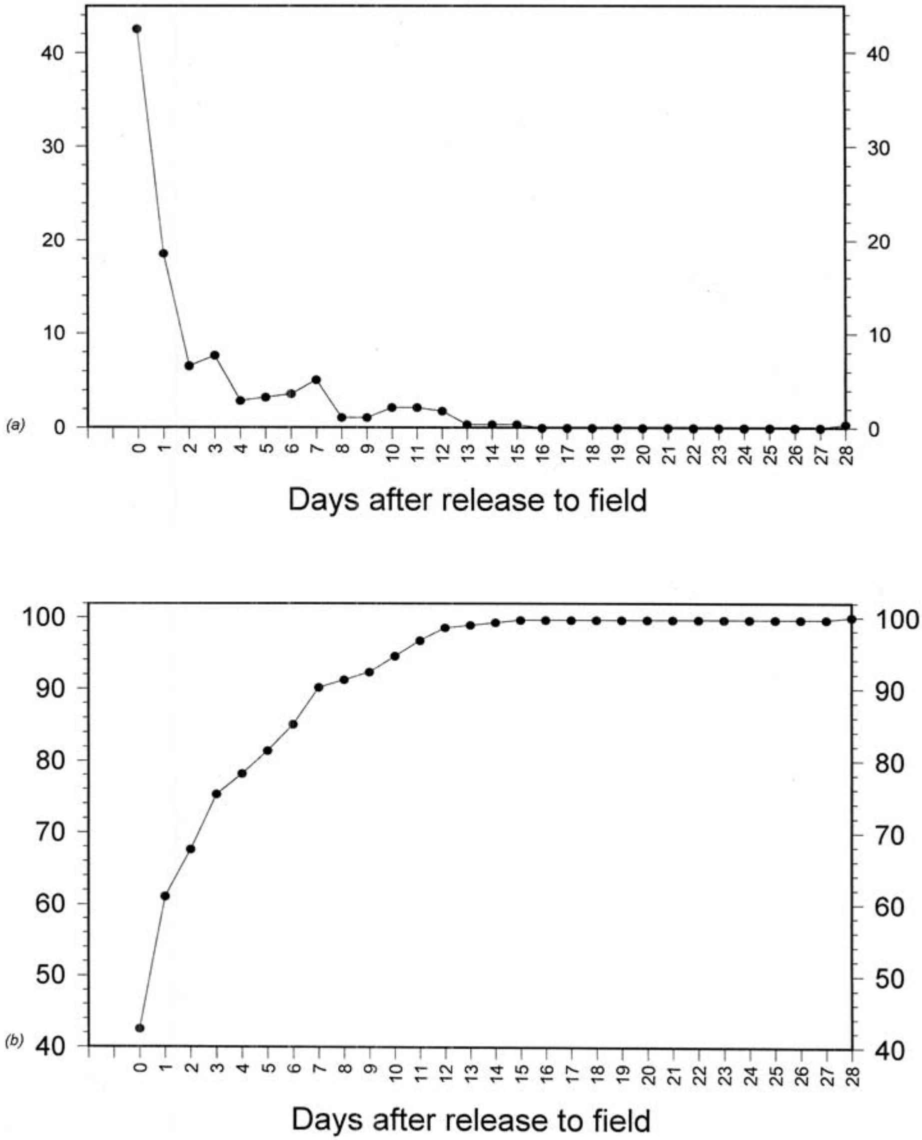

Fig. 2. Calls to completion, replicates released on July 5. (*a*) Percentage by day. (*b*) Cumulative percentage. (Data from 2000 Annenberg Election Survey.)

day, such that by the end of this day over 60 percent of interviews that would ultimately be completed were in the bank. Thereafter, increments were small: under 10 percent of ultimate completions for days 3 and 4 and under 5 percent for all succeeding days. By one week out, over 90 percent of ultimate completions had been recorded, and by two weeks (the nominal end of the interviewing window for any replicate) over 99 percent of interviews were in the bank. As it happens, one interview from the July 5 sample was conducted four weeks after release.

Now imagine the transposition of this sequence into the completion pattern for replicates released on later days. If the second day's replicates have exactly the same distribution as the first, then about 115 to 120 interviews on day 2 will be from that day's replicates, and about 50 interviews completed that day will be from replicates released the day before. The total number of interviews on day 2 should be about 165. On the third day, another 115 to 120 interviews will accrue from that day's release, along with about 50 from day 2 and 20 from day 1, for a total of about 210. The daily total will build for about two weeks, at which point earlier replicates will have been exhausted and dropped. From this point on, we can say that the day on which a respondent is interviewed is the product of a random draw.[4] Practically speaking, this is effectively true after about a week of interviewing so that from that time on the group of people interviewed on each day can be treated as a representative cross-section of the population.

Reality is slightly messier, but only slightly, according to figure 3. This figure tracks actual completions from July 5 to Election Day. The actual number of completions on July 5 is larger than the number implied in figure 2, as the NAES was already in the field. Before the Independence Day holiday, only one replicate was released per day, with an average daily completion rate of 50 interviews. The ramping up of fieldwork required close to two weeks, although the presence of open numbers from before July 5 accelerated the uptake modestly. In any case, from mid-July on, completions oscillated around the 300-person target.[5]

Advantages

For all the apparent complication of sample release and clearance, what results is just a set of daily cross-sections that can be combined into a large cross-section. And almost any temporal subsample can be combined into a

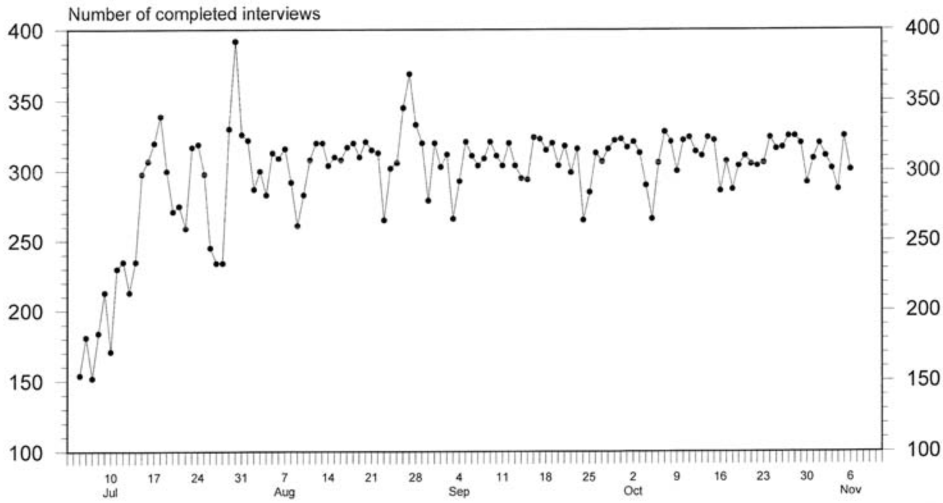

Fig. 3. Completed interviews by day, July 5 to Election Day. (Data from 2000 Annenberg Election Survey.)

representative cross-section, for example, the predebate, between-debate, and postdebate samples used in figure 1.[6] From this flow several benefits: low cost, uncontaminated respondents, and flexibility in choice of subperiods, with potentially high "granularity." Some of these advantages also carry corresponding disadvantages.

Low Cost

The cost of a rolling cross-section is only marginally more than of any other telephone cross-section with correspondingly aggressive clearance. Somewhat more management overhead is dictated by the need to monitor the process every day. A nonmonetary cost is paid in response rate, for two reasons. Assignment of the house's best interviewers to reluctant cases takes place earlier in a schedule that keeps cases open for only two weeks. This probably yields fewer conversions.[7] Most critical, however, is the arithmetic of the late campaign. Note in figure 3 that the rate of completion remained quite steady right to the end. Consider what this implies from the logic of figure 2. The replicates released on the last day of interviewing had one day only, that last day of interviewing, to be cleared. Replicates released the day before had only two days and so on. Inescapably, the NAES response rate started to drop two weeks before the end, and the same will be true, mutatis mutandis, for any rolling cross-section.[8] Initial response-rate decline is tiny, but it accelerates, essentially, on the temporal inverse of the pattern in figure 2b. The crucial point, however, is that each daily cross-section is representative of the population.

Fresh Respondents

Merely by virtue of their selection for a conversation that represents a more pointed focus on politics than most persons would ever experience in ordinary life, respondents are transformed by their encounter with a survey. But it seems likely that this conditioning is the result of the interview process, not just the fact of being contacted to do an interview. Consequently, compared to respondents at second and subsequent waves of panels, rolling cross-section respondents are only minimally conditioned. This seems like a major advantage even for events around which panels are commonly constructed. If one observes a pre–post effect from a debate, how much does that effect reflect conditioning from the initial interview?[9]

The price, of course, is that individual-level change cannot be captured. Analysis can close in on archetypal subgroups or profiles, to the extent that the variables defining group or profile membership are exogenous and thus insensitive to the passage of time. Demographic characteristics fit this criterion. Certain political "fundamentals" (Zaller 1998), such as party identification and ideological self-designation, come close but do not absolutely qualify.[10] Variables that respond to the campaign, such as exposure to and interest in media coverage, are clearly inappropriate as controls. And even the most stable of attribute distributions—race and religion, for example—still leave us short of capturing individual response.

Flexible Subperiods and "Granularity"

Figure 1 illustrates the flexibility of the design, as subperiods were adjusted after the fact to the dates of debates. In this case, we *chose* to divide the sample at the debates; we were not forced to do so by the logic of ex ante anticipation. At the same time, we achieve an efficient record of the time path of debate (or any event) effects, in that the first day after the event—as is also true of all further days—is just as much a cross-section as any single day or any number of days before the event. If the researcher is interested in, say, the immediacy with which impact unfolds and or in differences in time path between critical partitions of the sample (for example, whether or not the respondent saw any of the debate or the respondent's general exposure and attention to the mass media), relevant data not compromised by accessibility bias appear immediately after the event.[11] Meanwhile, joining up any combination of consecutive days is unproblematic. From the sampling perspective, all that adding or subtracting a day does is reduce or expand standard error. No other aspect of sample selection is being tweaked in the slightest. So the design combines flexibility in periodization with power in combination of days.

If this approach to seeking campaign effects seems ad hoc, the accusation is not troubling given the lack of theory about what drives the twists and turns of campaigns. Theory leaves us short of expectations for the identification of important campaign events, much less their timing and temporal shape. Holbrook (1996, chap. 6) very usefully and imaginatively goes beyond conventions and debates by considering other kinds of campaign events, but his criteria for selection events is "admittedly vague" (127), and he does not consider their temporal shape. Shaw's 1999 work catalogs alternative time paths suggested by control theory, and he tries to see what kinds of events follow which paths. But even he does not supply more than a typology of dependent-variable distributions. There is, so far, no body of propositions that might distinguish debates from conventions, for example, or a news impact from an advertising one. At the very least, the rolling cross-section allows us to maximize the variance to be explained. Down the road, it should allow us to build an inventory of dynamic patterns.

What makes this possible, of course, is daily management of data collection, such that any single day yields a random subsample of the total sample. Apart from the flexibility this affords us, the natural temptation is to focus on individual days, but the idea, just fronted, of identifying dynamic patterns involves more than juxtaposing consecutive days. It requires examining the pattern *within* that body of days or consecutive groups of small numbers of days, ideally by day-by-day comparison. The problem is illustrated by the example this essay opened with. We noted that there *did* seem to be some effect of the first debate, in that the immediately following days seemed to yield low readings, relative to the days before. But the days before also seemed to exhibit a drop in Gore's rating. This cast some doubt on the independent effect of the debate. But identifying an earlier point of discontinuity defied the naked eye. Realizing the full value of the design requires some mode of induction that captures the signal of true turning points from the noise of sampling error.

Graphical Smoothing

The Smoothing Problem

Smoothing methods for rolling cross-sections make it easier to identify the shape of true responses when they may be obscured by noise from sampling error, but smoothing methods must deal with two fundamental problems in separating "signal" from "noise." First, the shape of the signal poses analytical difficulties because different methods are needed to identify smooth versus abrupt changes in true responses. Yet, even for presidential debates, which have been studied for over forty years, we do not know whether poor performances have immediate impacts on public opinion like falling off a cliff or slower impacts like descending in an airplane. Second, sampling error presents problems because it creates noise that can be mistaken for real effects. Sampling error results from the variation in responses from observation to observation, and it is reduced by increasing the number of observations to allow the law of large numbers to smooth things out.

In election surveys, the variation in responses from subject to subject consists of two things: true differences in opinions across respondents and temporary differences within a respondent due to the vagaries of each person's response to questions. The former reflects heterogeneity in the population, and the later reflects measurement error due to imperfect questions and imperfect interviewing methods. If the population were homogeneous and every subject had the same opinion, then just one subject could be chosen to represent the population, but there is typically substantial heterogeneity in the population, and different people have different true opinions leading to variation from subject to subject. If we had a panel design in which the same person was observed again and again over time, then, putting aside the problems of conditioning and measurement error, any changes in the person's opinions could be ascribed to campaign events. But with heterogeneity and new cross-sections in every time period, variation in responses from one time period to the next could be due to differences in the people who are chosen and not the impacts of the campaign. The solution to this problem is to observe enough people in each time period—to have a large enough sample size—so that variation from one time period to the next due to sampling error is small compared to variation from campaign events.

Even if the population were homogenous or even if we had a panel, the variation from period to period might be due to measurement error—the different ways that people can answer questions when they have the same opinions. People with the same opinions may interpret the question in different ways because of interviewer effects or simply the imprecision of the question. Once again, with rolling cross-sections, the solution to this problem is to have a large enough sample size so that average measurement error is small compared to variation due to campaign events.

The Mean-Squared Error Criterion for Smoothing

A large sample size is the best way to smooth the data because it improves the signal to noise ratio by diminishing sampling error, but large samples are costly and constrained by limited budgets. Given the limits on daily sample sizes, the problem is to find the best way to extract signal from noise given the data at hand. For the Annenberg study this means finding the best way to analyze data on about three hundred respondents per day as shown in figure 3. The goal is to get the best rendition of the course of public opinion—the shape of the curve u_t where t is time and u_t is the true daily mean of opinion.

To do this, we need some criterion by which we can judge whether we have done a good or a bad job of smoothing the data. One criterion is unbiasedness. By this standard, if we want to know the population's average estimate of Gore's honesty (denoted by u_1, u_2, and u_3) for the three periods 1, 2, and 3, then the best estimate for each mean is the sample average of Gore's honesty ratings in each period from the three hundred respondents who were interviewed during that period. We denote these sample averages by u_1^*, u_2^*, and u_3^*. It is a standard result that with random sampling the expected value of each of these is equal to the true mean for that particular day; that is, u_1^*, u_2^*, and u_3^* are unbiased estimates of u_1, u_2, and u_3, respectively. For period 1, for example, this means that if we were to repeatedly sample from the population and get many estimates of u_1^* (say, from different polling firms operating on that same day), then the average of these many estimates would equal the true population value u_1. Unbiasedness of this sort is an especially useful property if we are looking for turning points because we want to make sure that each daily estimate is an unbiased estimate of the true signal for that day.

But another criterion is minimizing variance so that the standard errors of estimates of views about Gore's honesty are as small as possible. Since standard errors are a measurement of the amount by which our estimates vary from sample to sample, minimizing them means that we have reduced noise to a minimum. If we assume that the total variance due to heterogeneity and measurement error is s^2, then the sampling variance for each of u_1^*, u_2^*, and u_3^* is s^2/n, where $n = 300$, the number of observations in each period. For small n the quantity s^2/n can still be quite large. We could get an even smaller sampling variance if we assumed that Gore's honesty did not change over periods 1, 2, and 3 so that we could average u_1^*, u_2^*, and u_3^* to get $u^\# = (u_1^* + u_2^* + u_3^*)/3$ with a variance of $s^2/3n$—one-third the size of the previous sampling variance.

The quantity $u^\#$ is a three-period average, and if for any time-series with observations $t - 1$, t, and $t + 1$ we define $u_t^\# = (u_{t-1}^* + u_t^* + u_{t+1}^*)/3$, then we have a *three-period moving average.* The equal weights of (1/3, 1/3, 1/3) that define this estimator $u_t^\#$ are called the kernel weights in the statistical literature or just the kernel. Note that the weights always sum to one, but different patterns of weights define different estimators. The kernel for the estimator that takes just the *current period's mean* u_t^* from among $(u_{t-1}^*, u_t^*, u_{t+1})$ is (0,1,0). Hence, the kernel defines different ways to combine the daily sample means to produce an estimator, and kernels have different shapes ranging from the flat or "uniform" distribution with equal weights for the three-period moving average to the sharply peaked (at the middle value) shape for the current period's mean.

Unfortunately, $u^\#$ might be a biased estimate of even u_2, the middle period's value for Gore's honesty if Gore's honesty varies by period.[12] Thus, there is a tradeoff between bias and sampling variance, but we might be willing to trade a little bit of bias for a lot smaller sampling variance. There are, of course, limits to this, and we would not be willing to trade a lot of bias for a slightly reduced sampling variance. Statisticians formalize this tradeoff by considering mean squared error as a summary criterion for any estimator. Mean squared error is equal to the bias squared plus the sampling variance. To simplify the computation of mean squared error in this case, we set the zero point of the honesty scale by assuming that the middle value u_2 for honesty is equal to zero. There is no loss in generality in doing this because the scale is arbitrary to begin with. With this costless simplification, we can easily compute the mean squared error for $u^{\#}$ as an estimator of u_2:[13]

$$\mathrm{MSE}(u^{\#}) = [u_1 + u_3]^2/9 + s^2/3n.$$

The formula for the bias term $[u_1 + u_3]^2/9$ may not seem intuitively obvious so we explore its properties in more detail later. We can also compute the mean squared error from using u_2^* as the estimator for u_2, which will be just equal to the variance of u_2^* since u_2^* is an unbiased estimator of u_2:

$$\mathrm{MSE}(u_2^*) = s^2/n.$$

Under what conditions should we use $u^{\#}$ versus u_2^*? That is, under what conditions should we use the three-period moving average versus the one period estimate? As the variance s^2 gets bigger, there is more noise in the data. It makes sense in this situation to use $u^{\#}$ instead of u_2^* because the bias term in $\mathrm{MSE}(u^{\#})$, that is, $[u_1 + u_3]^2/9$, will be dominated by the variance term, $s^2/3n$, so it is worth accepting some bias from averaging all three periods to get the much smaller variance term ($s^2/3n$) in $u^{\#}$ compared to that (s^2/n) in u_2^*. As n gets bigger and bigger, the bias term in $\mathrm{MSE}(u^{\#})$ will dominate the variance term so it will make sense to use u_2^*, which does not have any bias term. With more observations, there is no need to average over adjoining periods in order to reduce noise—the number of observations in a single period does that nicely. Similarly, as the bias term gets bigger and bigger, it makes sense to use u_2^* instead of $u^{\#}$ because u_2^* is unbiased. That is, if the twists and turns of the campaign cause the variable of interest to change a lot, then we should refrain from averaging over adjacent periods.[14] In summary, for large variance, small n, and small bias, it makes sense to use $u^{\#}$. For small variance, large n, and large bias, we should use u_2^*.

The Shape of the Response Curve as a Crucial Factor

The bias term, $(u_1 + u_3)^2$, deserves some additional discussion because it summarizes the shape of the response curve—the way that ratings of Gore's honesty can be expected to go up and down. Because we have set $u_2 = 0$, this quantity attains its smallest possible value of zero when u_1 and u_3 are also zero. In this case, it clearly makes sense to combine the three periods to estimate Gore's honesty because the true value is the same for all three periods. But the bias term also attains its smallest possible value when $u_1 = -u_3$ and the three points lie along a straight line u_t with a constant slope. Thus, the bias term is only nonzero when the three points u_1, u_2, and u_3 depart from lying along a straight line—only when the slope of the curve u_t is changing. The classic measure of the change in a slope is the second derivative of the curve. Consider the standard difference-in-differences approximation of a second derivative:

$$u_t'' = \partial^2 u(t)/\partial t^2 \approx \{[(u_3 - u_2)/h] - [(u_2 - u_1)/h]\}/h,$$

where h is some small unit of time. Since we have assumed that $u_2 = 0$, this amounts to $u_t'' \approx [u_3 + u_1]/h^2$ so that by a little algebra, $h^2\, u_t'' \approx [u_3 + u_1]$, which is the square root of the bias term. This result will come in handy later because $[u_t'']^2$ is a convenient measure of the shape of the response that we want to detect.

Choosing the Bandwidth

So far, we have been thinking about the problem of smoothing as one in which we want to choose the optimal weights (or kernel) for an estimator that uses data from three time periods. In some cases, as discussed previously, this might mean that very little smoothing takes place, and in others it might mean that a great deal of smoothing occurs. There is, however, another way to think of the problem. Instead of choosing the optimal kernel for three periods, we might choose a particular kernel shape (uniform, peaked, or some other shape) and then ask how many periods should be smoothed with this kernel. If the number of periods is very small (for example, a few hours instead of weeks or months, which consist of many hours), then the kernel will smooth over a very short period of time. If the number of periods is very large, then the kernel will smooth over a longer period of time. One reason for reformulating the problem in this way is that it turns out that the shape of the kernel typically matters less than the period of time over which the smoothing takes place, and there appears to be good arguments for always choosing particular shapes such as the parabolic Epanechnikov kernel (Hardle 1990, 24–26, 133–37).

To reformulate the problem in this way, suppose that observations are spread evenly over the total campaign that is being studied so that we can slice the time periods smaller and smaller and still get some observations. Thus months can be split into weeks and weeks into days. Of course, there is a limit to how far we can do this with the rolling cross-section design because our smallest unit is a day, but this thought experiment is nevertheless useful. Assume that the total length of the time period on the horizontal axis is one unit and that there are N observations spread evenly over this entire time period. We break the horizontal axis into a number of evenly spaced time periods, each of which is h units apart, where h is some fraction of one. These time units might be months, weeks, or days. Then for any given time period, there are $n = hN$ observations. Our goal will be to see what happens as we change h, which is called the bandwidth in smoothing language. Intuitively, we would expect that as the bandwidth h gets smaller, the amount of bias in the estimator will decrease, but the number of observations $n = hN$ will also get smaller, causing the variance in the estimator to increase. Thus the choice of bandwidth is an essential aspect of choosing a smoother because a good choice will minimize the mean squared error.

We choose the equal weighting kernel (moving average) so that the mean squared error is as stated earlier:

$$\mathrm{MSE}(u^{\#}) = [u_1 + u_3]^2/9 + s^2/3n.$$

Using the previous result for the bias $[h^2\, u_t'']\approx[u_3 + u_1]$ and the fact that $n = hN$, we can rewrite this as[15]

$$\mathrm{MSE}(u^{\#}) \approx h^4[u_t'']^2/9 + s^2/3hN.$$

In words, we can write the MSE as follows:[16]

$$\mathrm{MSE}(u^{\#}) \approx (\text{Bandwidth})^4\ (\text{Deviation from linearity of curve})^2/9$$
$$+\ [\text{Variance of Error}]/[3 \times \text{Bandwidth} \times \text{Total Sample Size}]$$

Just as we expected, as the bandwidth h gets bigger, the bias term increases but the variance term gets smaller. Furthermore, the amount of bias depends upon the size of the second derivative and the curve's deviation from linearity. The more "wiggly" the curve, the more bias there is in the estimator.

To minimize mean squared error, we can choose the optimal bandwidth by taking the derivative of MSE($u^{\#}$) with respect to h, setting the resulting derivative to zero, and solving for h. We obtain

$$h^5 = 3s^2/[4N(u_t'')^2].$$

The optimal bandwidth gets wider with greater population variance s^2 and narrower with increasing N and increasingly wiggly curves.

Analyzing the Annenberg Data on Gore's Honesty

We can use this formula to determine the optimal smoothing for the Annenberg data on Gore's honesty. From the daily data, we can estimate s^2 as the average of the daily variances.[17] The result is a value of about .48. The value of N for the sixty-eight days of interviewing is 20,892. The value of u_t'' depends upon the kinds of responses we expect to find in the general population. One natural measure of a unit of response is the cross-sectional standard deviation in the variable of interest, such as perceptions of Gore's honesty. Changes in the mean equivalent to about 5 percent of the standard deviation might be considered significant, although they might also be hard to detect. Changes in the mean equivalent to about 25 percent of the standard deviation would certainly be substantial, and we would want to be able to detect them. Consider each of these possibilities.

Assume that we expect changes in the mean value of Gore's honesty of one-quarter of the cross-sectional standard deviation, and assume that we expect that these changes might happen within three days. Then we can calculate an approximate value for u_t'' as follows. Suppose that the trend line is flat and that it changes upward (or downward) by one-quarter of a standard deviation (.25 units on the honesty scale) in three days, which is about one-twentieth (.05) of the total sixty-eight days on figures 1a and 1b. Then u_t'' will be $.25/.05 = 5$ over this period. Putting this number along with the variance s^2 (.48) and the total number of interviews (20,892) in the previous formula yields $h = .058$ so that $n = hN$ will be about 1,200, or four days of interviewing at three hundred respondents per day. Similarly, if we are expecting changes of 5 percent of a standard deviation, then $u_t'' = .05/.05 = 1$ and $h = .111$ so that n will be about 2,400, or eight days of interviewing. These results suggest that the ideal amount of smoothing will be something like four to eight days.

Figure 4 presents three-, five-, and seven-day moving averages of the data on perceptions of Gore's honesty presented in figure 1.[18] The most remarkable feature of these graphs is the strong impression that Gore's decline started well before the first debate—perhaps as much as two weeks before and certainly ten days before. The debate may have accelerated the decay in judgment on Gore. Certainly the downward slope seems to increase its pitch right after the debate. Then again, the total drop after the debate is no greater than what occurred before. But the basic point stands: the turning point that is the pivot for the whole campaign came two weeks *before* the debates. This fact would almost certainly have been missed by any other design for fieldwork. Johnston, Hagen, and Jamieson (2004, chap. 6) interpret that turning point in terms of media attacks on Gore's credibility, starting with stories about factual errors in claims made in high-profile speeches and ending with the controversy over his request to President Clinton to release oil from the nation's strategic reserve. They would not have been led to any such interpretation but for the rolling cross-section design.

This discussion has just scratched the surface of what can be done with smoothing methods, and it has used one of the very simplest methods. A several day moving average is a very simple form of the more general technique of polynomial smoothing where the data at each point are approximated by a weighted polynomial regression. In the case of moving averages, this "regression" is just a constant produced by the weighted average of nearby observations where the kernel defines the weights and the bandwidth defines what is considered nearby. More sophisticated polynomial smoothing methods fit local linear regressions (with a constant and a linear time term) or higher order polynomials to each point by estimating a polynomial regression around that point with weights equal to the kernel weights.

Since the publication of William Cleveland's "Robust Locally Weighted Regression and Smoothing Scatterplots" (1979), most researchers favor polynomial methods that at least use linear regressions and that employ robust methods that reduce the impacts of local outliers. Cleveland's LOWESS (Locally weighted scatterplot smoothing) or LOESS is applied to the data on Gore's honesty in figure 5 with different bandwidths ranging from .07 to .125. The results are similar to the moving averages in figure 4.

There are also many smoothing methods other than polynomial smoothing, of which the most popular is various forms of splines (see

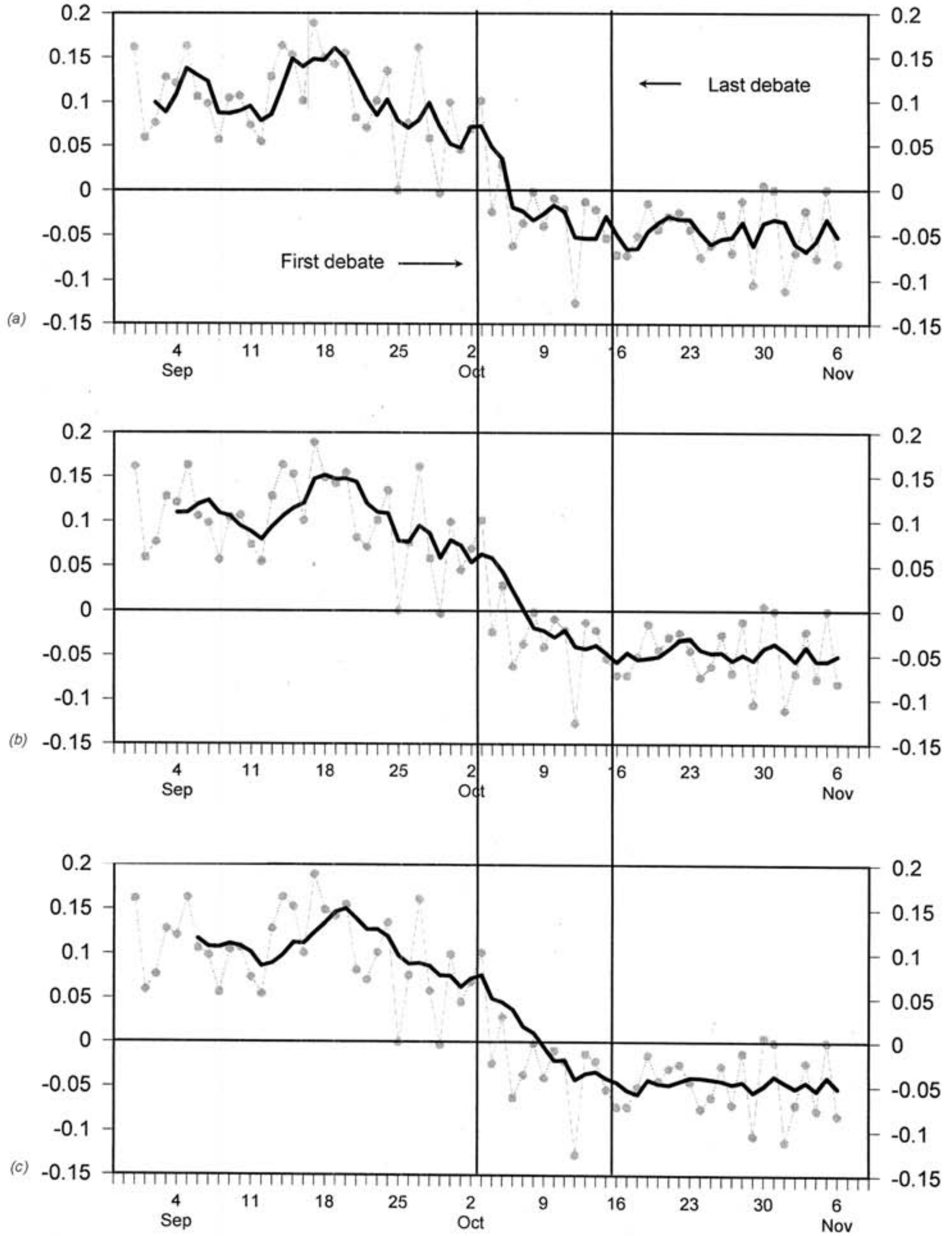

Fig. 4. Smoothing by prior moving average. (*a*) Three-day. (*b*) Five-day. (*c*) Seven-day.

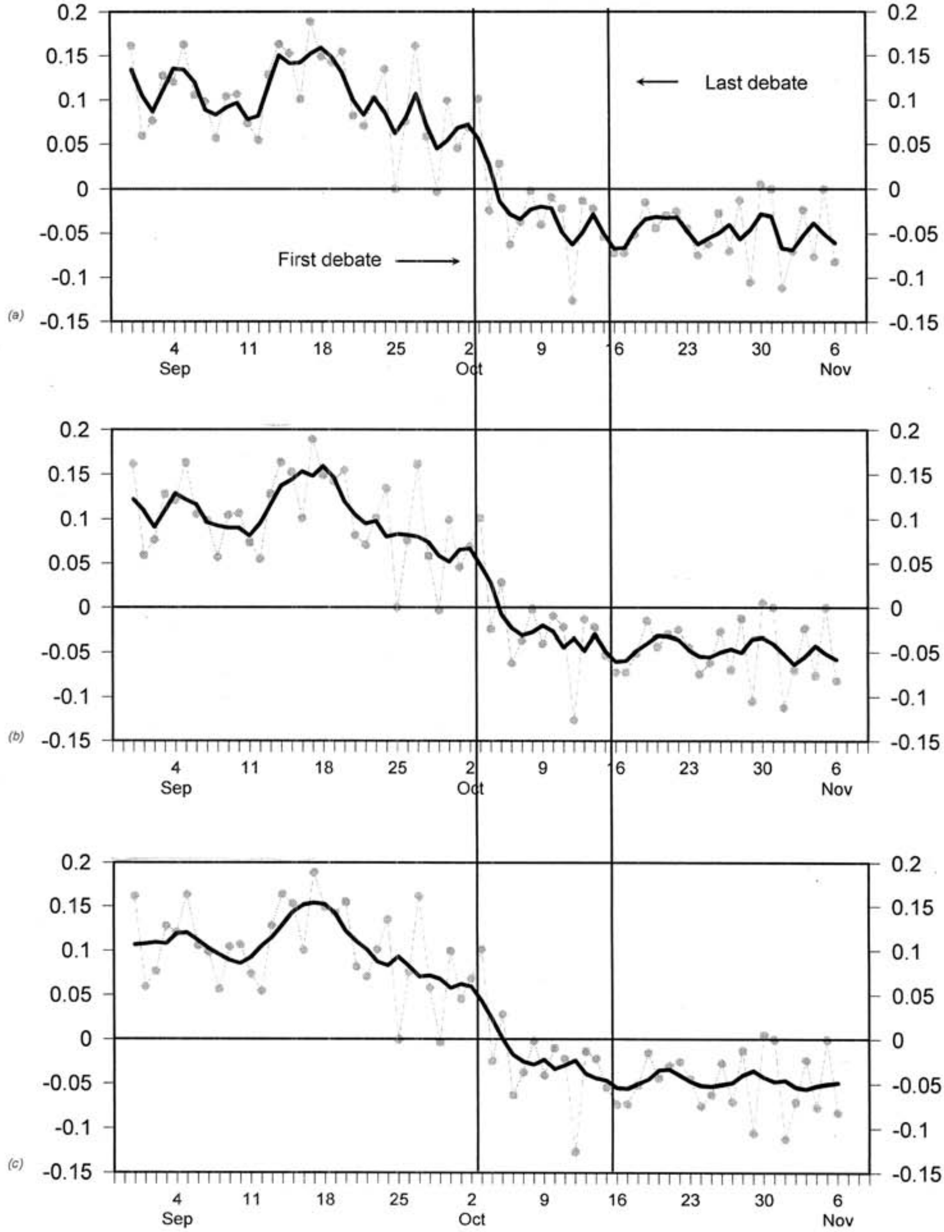

Fig. 5. Smoothing by LOESS. (*a*) bandwidth = 0.075. (*b*) bandwidth = 0.100. (*c*) bandwidth = 0.125.

Green and Silverman 1994; Ruppert, Wand, and Carroll 2003) that smooth data by fitting them to piecewise polynomial (often linear) functions that are spliced together at knots. There are close relationships among these methods. Silverman (1985, 3–4), for example, shows that spline smoothing can be considered a form of weighted moving average smoothing with a particular kernel and varying bandwidth (see also Hardle 1990, 56–64).

Although we have focused on the optimal smoothing problem, statisticians have also given considerable attention to the problems of inference from smoothed data, and they have developed methods for describing confidence intervals for the curves produced by smoothing. These methods provide ways to address the statistical power issues highlighted by John Zaller (2002) in his studies of the inferences that can be made from election studies. Some representative references are Hardle 1990, chapter 4; and Ruppert, Wand, and Carroll 2003, chapter 6.

A Mixed Design

Although this essay began by treating the rolling cross-section and the panel design as substitutes, in fact they are better seen as complements. That is, a properly constructed election survey can be both a rolling-cross section and a panel; all that is required is for one wave of interviewing to have controlled release of sample. It might be tempting to deploy a precampaign cross-section as a baseline and then meter the next wave over the campaign. But if the point of the initial wave is to establish a baseline, this can be achieved by examination of distributions and parameters in the early days of the campaign. As most comparisons to baseline that truly matter are aggregate ones, reinterviewing adds little value. Meanwhile, panel conditioning might distort estimates of aggregate change. It might be best to compare fresh cross-section with fresh cross-section.

The obvious way to connect the designs is with a simple pre–post setup, where the preelection, or campaign, wave is the temporally metered rolling cross-section. Temporal metering of the postelection wave might also be undertaken but on a different basis: release of numbers at the postelection wave should be as uncorrelated as possible to the timing of first-wave completions.[19] This makes it possible to do two things. First, the postelection rolling cross-sections can be used to monitor aggregate changes that occur in the postelection period in the same way as the preelection rolling cross-sections are used. Second, over-time changes in individuals can be more reliably ascribed to events if there is no correlation between the interview date on the postelection wave and the preelection rolling cross-sections.

Suppose, for example, that we wanted to know whether a debate had a negative impact on the rating of a candidate. The obvious approach would be to take people interviewed before the debate and those interviewed after the debate and to calculate the difference in their preelection rating and their postelection rating of the candidate. If this difference is smaller for the postdebate group than for the predebate group, then we might claim that the debate did reduce the candidate's rating. But this claim would be at risk if there was a correlation between the preelection interview date and the postelection interview date. In the worst scenario, all those interviewed before the debate might have been reinterviewed within two weeks of the election while those initially interviewed after the debate might have been reinterviewed several weeks after the election. Then some negative event two weeks into the *post*election period might explain the difference-in-differences that was found. This possibility can be ruled out by making sure that the date of reinterview is uncorrelated with the initial interview date.[20]

When done in this way, combining designs increases the power of each. The panel element benefits from the fine granularity of intervals between the pre- and the postelection interview, which makes it possible to determine how time and events affect opinion change. Moreover, the panel design is greatly strengthened by paying attention to this granularity and making sure that the date of the initial interview and the reinterview are uncorrelated. Indeed, it is quite possible that panels have sometimes suffered from correlations between the two dates that may have confounded inferences made from them.

At the same time, there is no penalty in the more prosaic, but no less useful, features of the pre–post design. For example, questions where postelection rationalization is a problem because voters might try to make their opinions fit their electoral choices (or might try to make their opinions fit the electoral choices of the majority of voters) are best asked before the election when final choices have not yet been made, even if they are subsequently used in analysis of questions such as vote choice best asked after the event. There will, of course, be temporal heterogeneity in the preelection response, with potentially adverse impact on the stability of estimates. But this takes us right back to where we started. *Any* fieldwork that stretches over more than a few days is likely to have temporal heterogeneity, as several essays in this volume show. It is best that that heterogeneity be recognized explicitly, guarded against if possible by making sure that the dates of interviews are uncorrelated with one another, and ultimately modeled directly if it still remains a problem. And, of course, the heterogeneity produced by events ought not to be confused with heterogeneity produced by differences in respondent accessibility.

The rolling cross-section component benefits from the merger of the two designs by clearer separation of cross-sectional and longitudinal variance. A simple example of such leverage is portrayed in figure 6, for analysis of a debate effect. If one observes after a debate a difference between those who saw the debate and those who did not, is the difference the result of actual exposure to the debate, or is it merely symptomatic of an abiding difference that also correlates with the likelihood of viewing the debate in the first place? By itself, a rolling cross-section data file cannot address this question. But one linked to a postelection wave can. A critical fact about the postelection wave is that debate exposure information can be gleaned from *all* respondents, including those first interviewed before the debate. The postelection data allow us to read back through the event and to distinguish its endogenous and exogenous components.

The example in figure 6 is from the 1988 Canadian Election Study. In that year's debate among the party leaders, John Turner of the Liberal Party apparently scored a clear victory. This both primed and moved opinion on the main issue, Canadian–U.S. free trade, and it rehabilitated Turner's reputation as a leader. Figure 6 shows the extent to which this rehabilitation was conditional on exposure to the event itself. Exposure is indicated by response to a postelection question, and so the comparison extends back virtually to the start of the campaign. Smoothing is by prior moving averages (the technique exemplified in fig. 4 and in Johnston, Hagen, and Jamieson 2004), and so any turning points should be correctly located. There is the merest hint that respondents who would watch the debate began to reevaluate Turner just before the event.[21] In general, however, it appears that debate watchers did not bring any different beliefs to the moment than did nonwatchers. So the difference right after the event is mostly real, in the sense that it truly reflects impact from the moment, not from selection bias.

Impact from the moment is not entirely the same as impact from the debate, however. This potential indeterminacy shows how leverage can work the other way.

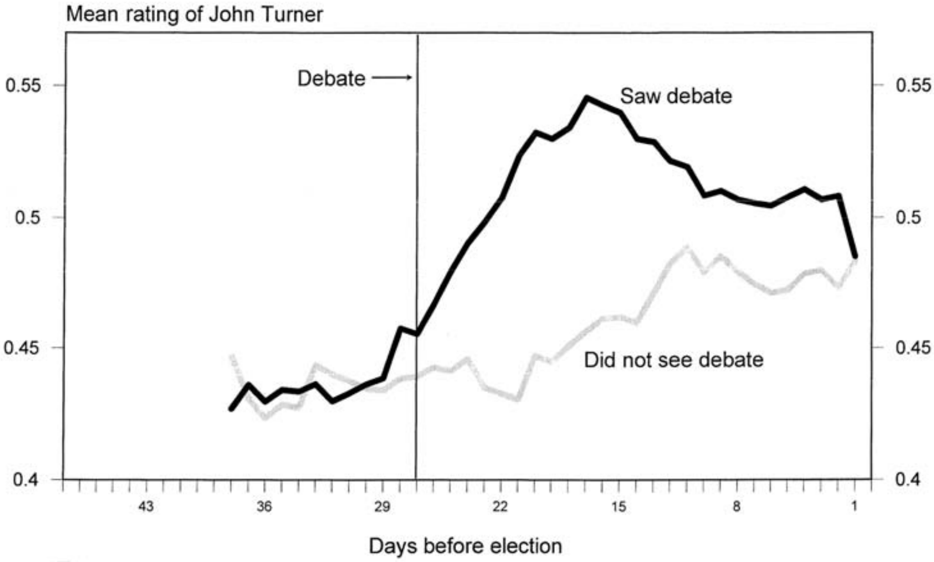

Fig. 6. Impact of debates (English-speaking respondents only), smoothed by nine-day prior moving average. (Data from 1988 Canadian Election Study.)

Even if watchers and nonwatchers bring identical priors to the moment, the difference that subsequently emerges between the groups may not be from direct exposure to the event as opposed to coverage of it. Inspection of daily values suggests that most of the postdebate gap appeared the day after the debate, and so it is probably the result of debate viewing itself, not of general media exposure. But the gap continued to grow for a few days, and this is unlikely to be just the product of unassisted reflection. Johnston et al. 1992 show that the same media orientation that delivered an audience for the debate itself also produced exposure to the generally positive coverage of Turner that followed the debate.[22] A simple pre–post design, even if the second wave follows immediately on the event, would struggle to separate these components.

Late in the campaign, coverage of Turner turned negative, and that fact is also reflected in the data. Respondents who saw the debate were also sensitive to this shift and turned against the Liberal leader, such that the net effect of the whole campaign on evaluation of Turner was very modest. Nonwatchers, meanwhile, absorbed campaign stimuli, only with some lag. By the end, watcher/nonwatcher differences were completely washed away. This shows that conditioning on a debate exposure question asked more than a few days after the event is likely to yield to a false negative for the event.[23]

Multivariate Analysis of the Rolling Cross-Section Design

The simple bivariate example in figure 6 belongs to a more general class of cases. Many relationships in a rolling cross-section data set embody both cross-sectional and longitudinal covariance—cross-sectional because people have different reactions to the same events and longitudinal because people are exposed to different events over time (Johnston and Brady 2002; Zaller 2002). Ignoring time when estimating a cross-sectional relationship or ignoring cross-sectional variation when estimating a time-series risks conflating the two components. The most obvious problem is the failure to specify the right relationship when either time or individual variation is ignored. But attitudinal data present special problems that can confound even sensible specifications. Many variables that shift over a campaign and so have potential dynamic significance also carry "projective" cross-sectional variance from, for example, respondents' party identifications. Perceptions of honesty, for example, not only change with events, but they are also the result of party identification, which colors people's perceptions about the candidates. To an extent, this projective component can be mitigated by controlling for the variable that is its source—by, for example, analyzing people with different party identifications separately or including party identification in a regression equation. But measurement error in party identification and other individual characteristics may mean that control is incomplete so that covariance is, as it were, left on the table for the target variable of interest (such as honesty) to pick up—leading to false inferences about the impact of the target variable.

In the ideal case, we should isolate the target variable's longitudinal component, the element least likely to carry extraneous impact, from cross-sectional variation, but this is not always easy to do. Johnston and Brady (2002) include an extensive discussion of how to approach the problem. They make a case that the ideal way to proceed is to exploit the study's panel property, if one exists. This requires a pre–post design with repetition (where possible) of questions asked in the campaign wave. In a properly conducted postelection wave, variance on the key indicators should be only cross-sectional. To this end, postelection fieldwork should be conducted with deliberate speed. To the extent that fieldwork is stretched out, rerelease of sample should be, as described previously, uncorrelated with the campaign-period release schedule. Thus, even if time lurks in the postelection data, it is essentially irrelevant to impact from true "campaign time." With such data in hand, the postelection wave captures the cross-sectional component in any relationship. By including these postelection indicators in the estimation, impact from the preelection indicators is rendered longitudinal.

What can be done, however, when only repeated cross-sections are available? Johnston and Brady (2002) also considered multivariate methods for analyzing rolling cross-sections when there is no panel information, but their method works best with large cross-sections. Deaton (1985) considers a related problem with a time-series of repeated cross-sections, and he shows how cohorts, defined as groups with fixed membership and exogenously defined characteristics, can be tracked through these data and corrections can be made for cohort-based fixed effects using errors-in-variables methods. Moffitt (1993) extends this analysis in several ways, such as the consideration of autoregressive linear models and discrete dependent variables. Verbeek and Nijman (1992a) consider the general question of whether cohort data can be considered as genuine panel data, and they summarize the state of knowledge about "Pseudo-Panels" in Verbeek and Nijman 1992b. Much more work needs to be done to adapt these methods to the circumstances of election studies where measurements contain substantial error and where the cross-sections are very small and there are many temporal observations.

Conclusions

The rolling cross-section design is a powerful one for detecting the impact of events over time, and it has strengths that are lacking in the standard panel design. Indeed, we show that panels can miss important turning points and events. Moreover, designers of panels have probably underestimated the problems that can arise in making inferences from panels when fieldwork for each wave is spread out, as it almost inevitably has to be, over a period of time. Because of its focus on temporal change, the rolling cross-section design suggests ways that panels themselves could be improved by incorporating rolling cross-sections in each wave.

Despite its advantages, the rolling cross-section design also presents substantial analytical challenges due to the small size of each period's sample and the problem of separating out individual cross-sectional variability from temporal change. Panels provide some advantages with respect to this problem, but new statistical techniques have made it possible to attack these difficulties in fruitful ways for rolling cross-sections. Moreover, both practical experience and theoretical advances with the design suggest that a hybrid of the rolling cross-section with a culminating panel can provide substantial inferential power.

NOTES

1. For details on the NAES, see Romer et al. 2003.

2. The question is, "Does the word 'honest' describe Al Gore extremely well, quite well, not too well, or not well at all?"

3. The rigidity of the callback sequence is modified when interviewers pursue opportunities presented by the field. For example, if a contact expresses willingness to make an appointment outside the normal two-week clearance window, interviewers are commonly instructed to make the appointment, as maximizing response rate is always a serious priority. Similarly, if a call attempt indicates that the respondent is at home but already engaged with a call, the interviewer will phone back promptly.

4. Thus, although the completions on a given day come from different replicates—from today's, the preceding day's, the replicate from the day before that, and so forth—they should still amount to a random sample of the population if the samples have all been worked with the same intensity. By working the samples with the same intensity, we ensure that today's interviews from the replicate of five days ago are statistically valid substitutes for the group of people from today's replicate who will be ultimately interviewed five days from now. The "same intensity" assumption, therefore, allows us to make the jump from random replicates to the assumption that those interviewed on a given day represent a representative sample of the population.

5. The data show that the fieldwork house, Shulman, Ronca, and Bucuvalas, Inc. (SRBI), struggled early on to find the target but was on top of the task by early August.

6. The only exceptions to this rule are samples from transitional subperiods, such as the days following July 5. Relatively inaccessible respondents will be underrepresented, relative to other periods, at transitions that involve increasing sample size and overrepresented at transitions involving reductions in sample size. Analysis for transitional days should, strictly speaking, employ weights for accessibility.

7. We are grateful to David Northrup, project director on the Canadian Election Studies, for this insight.

8. Well, maybe not in Canada. In the 1993, 1997, and 2000 Canadian Election Surveys, completion numbers climb in the last week, quite without any change in fieldwork intensity. The heart of the matter seems to be that respondents who

earlier would schedule a later interview now agree on the spot or agree to be interviewed promptly, under the shadow of the deadline.

9. Some leverage on this question could be gained by drawing a fresh post-debate cross-section and using this for calibration. This starts to inflate costs, however, and it presents its own comparison problems, as the second wave of the panel is not itself a cross-section.

10. Johnston, Hagen, and Jamieson (2004), for example, find that both party identification and liberal/conservative ideology drift toward the temporarily advantaged party and then away as the advantage shifts. Such endogenous movement in party identification can be minimized by using response to the root question. This means that the seven-point scale, where "leaners" are assigned to parties and partisans assigned intensity scores, is inappropriate for rolling cross-section analysis.

11. Postevent surveys started right after the event and completed within a day or two overrepresent those respondents who are easily accessible by the interview method. The rolling cross-section overcomes this problem. Consider, for example, a population in which "stay-at-homes" almost always answer on the first day of interviewing whereas those who "get-out-of-the-house" typically require several days of calls. Further assume that after a week's effort, both groups are just about as likely to be interviewed. A postevent survey conducted for one or two days would have very high response rates for "stay-at-homes" and consist mostly of such people. A rolling cross-section would interview the correct proportions of each group because it would pick-up those who "get-out-of-the-house" and were not interviewed before the event from replicates released before the event. If "stay-at-homes" are different from those who "get-out-of-the-house" (and there is abundant evidence that they are), then the postevent survey will provide a biased picture of the impact of the event. Brady and Orren (1992) provide an example with respect to the Canadian debates.

12. The bias in using $u^{\#}$ to estimate u_2 comes from using information that is one period away from period 2 (namely, information from periods 1 and 3) as well as contemporaneous information. It seems likely that there will be even more bias in using $u^{\#}$ to estimate u_1 or u_3 because $u^{\#}$ uses some information from two periods away.

13. We can generalize the result a bit by assuming that $u^{\#} = a_1 u_1^* + a_2 u_2^* + a_3 u_3^*$ with the weights adding up to one ($a_1 + a_2 + a_3 = 1$). Since $u^{\#}$ is an estimator for u_2, it makes sense to assume a symmetrical treatment of period 1 and period 3 observations so that $a_1 = a_3$. Then we can write $u^{\#} = au_1^* + (1 - 2a)u_2^* + au_3^*$. The expected value of this is $E(u^{\#}) = au_1 + (1 - 2a)u_2 + au_3$, and the true value of the period 2 average is u_2 so that the expected bias is $Bias(u^{\#}) = E(u^{\#}) - u_2 = au_1 + (-2au_2) + au_3$. Since we have set $u_2 = 0$, this simplifies to $Bias(u^{\#}) = a(u_1 + u_3)$. The variance of $u^{\#}$ can also be easily calculated as $\mathrm{Var}(u^{\#}) = (6a^2 - 4a + 1)s^2/n$. Hence, the mean squared error is

$$\mathrm{MSE}(u^{\#}) = a^2(u_1 + u_3)^2 + (6a^2 - 4a + 1)s^2/n.$$

If $a = 1/3$, then this becomes the expression in the text for the three-period moving average.

14. This analysis can be done more formally with the results from the preced-

ing footnote by minimizing the mean squared error in that footnote with respect to the parameter a, which produces $u^{\#}$ when $a = 1/3$ and $u_2{}^*$ when $a = 0$. We can find the value of a by taking the derivative of the MSE in the preceding footnote with respect to a, setting the derivative equal to zero, and solving for a. The result, after some algebra, is

$$a = 2/\{6 + n[(u_1 + u_3)^2/s^2]\}.$$

Clearly this has the two limits 1/3 (producing $u^{\#}$) and zero (producing $u_2{}^*$). Furthermore, for small n or small $(u_1 + u_3)^2$, we obtain something close to $u^{\#}$, whereas for large n or large $(u_1 + u_3)^2$ we get $u_2{}^*$. For large s^2 we get $u^{\#}$, and for small s^2 we get $u_2{}^*$.

15. Note that we conveniently choose the interval h for computing the approximation to the derivative to be the same as the bandwidth. This means that the expression for MSE($u^{\#}$) is only approximate.

16. If we carry through with the more general case in the preceding footnotes, we get that

$$\text{MSE}(u^{\#}) \approx a^2\, h^4[u''(t)]^2 + [6a^2 - 4a + 1]\, s^2/hN.$$

And if we think of a kernel as a function $K(t)$ that defines weights for each value of t, then we can define

$$c = \textstyle\sum_t [K(t)]^2 = \text{Sum of Square of Kernel Weights} = 6a^2 - 4a + 1$$

$$d = \textstyle\sum_t u^2\, K(u) = \text{Variance of Kernel weights} = 2a,$$

so that we can write $\text{MSE}(u^{\#}) \approx h^4(d^2/4)[u''(t)]^2 + cs^2/hN$. This result is identical to the general result of Gasser and Muller reported in Hardle as Theorem 3.1.1 (Hardle 1990, 29–30).

17. We are eliding a potential complication here by assuming that s^2 is constant across the campaign. Brady and Johnston (1987, 170–73) show how standard deviations for trait batteries become greater over the course of a primary campaign (see also Campbell 2000; Wlezien and Erickson 2002). In this case, the variation in s^2 is probably a second-order problem, but it will not be in every case.

18. We use "prior" moving averages in which the smoothed point on day t from a p-period moving average is calculated from average of day $t, t - 1, t - 2, \ldots, t - p - 1$. Prior moving averages have the virtue that, if a turning point occurs in the underlying true series u_t, then the prior moving average will only start to turn at the point where the true series begins to turn. They have the defect that the prior moving average may underestimate the size of the turn.

19. It is impossible to ensure that the actual gap between first- and second-wave interviews is uncorrelated to first-wave timing. The closest we can come is to make rerelease of the number to field uncorrelated to the initial completion date.

20. Note that this analysis is symmetrical and that it would also allow inferences about the impacts of postelection events by comparing groups interviewed before and after some postelection occurrence.

21. The predebate uptick among eventual debate watchers reflects one outlier. All other observations for this group in the period indicate no predebate shift.

22. Indeed, much of this coverage was simple repetition of the key moment in the debate.

23. This observation applies to any postevent retrospective question, not just one posed in the second wave of a panel.

REFERENCES

Brady, Henry E., and Richard Johnston. 1987. "What's the Primary Message: Horse Race or Issue Journalism?" In *Media and Momentum,* ed. Gary Orren and Nelson Polsby. Chatham, NJ: Chatham House.

Brady, Henry E., and Gary Orren. 1992. "Polling Pitfalls: Sources of Error in Public Opinion Surveys." In *Media Polls in American Politics,* ed. Thomas Mann and Gary Orren. Washington, DC: Brookings Institution.

Campbell, James E. 2000. *The American Campaign: U.S. Presidential Campaigns and the National Vote.* College Station: Texas A&M Press.

Cleveland, William. 1979. "Robust Locally Weighted Regression and Smoothing Scatterplots." *Journal of the American Statistical Association* 74:829–36.

Deaton, Angus. 1985. "Panel Data from Time Series of Cross-Sections." *Journal of Econometrics* 30:109–26.

Dunkelberg, William C., and George S. Day. 1973. "Nonresponse Bias and Callbacks in Sample Surveys." *Journal of Marketing Research* 10:160–68.

Green, P. J., and B. W. Silverman. 1994. *Nonparametric Regression and Generalized Linear Models.* London: Chapman and Hall.

Groves, Robert M. 1989. *Survey Errors and Survey Costs.* New York: Wiley.

Hardle, Wolfgang. 1990. *Applied Nonparametric Regression.* Cambridge: Cambridge University Press.

Hawkins, Thomas M. 1975. "Estimation of Nonresponse Bias." *Sociological Methods and Research* 3:461–88.

Holbrook, Thomas M. 1996. *Do Campaigns Matter?* Thousand Oaks, CA: Sage.

Johnston, Richard, André Blais, Henry E. Brady, and Jean Crête. 1992. *Letting the People Decide: Dynamics of a Canadian Election.* Stanford: Stanford University Press.

Johnston, Richard, and Henry E. Brady. 2002. "The Rolling Cross-Section Design." *Electoral Studies* 21:283–95.

Johnston, Richard, Michael G. Hagen, and Kathleen Hall Jamieson. 2004. *The 2000 Presidential Election and the Foundations of Party Politics.* Cambridge: Cambridge University Press.

Moffitt, Robert. 1993. "Identification and Estimation of Dynamic Models with a Time-Series of Repeated Cross-Sections." *Journal of Econometrics* 59:99–123.

Romer, Daniel, Kate Kenski, Paul Waldman, Christopher Adasiewicz, and Kathleen Hall Jamieson. 2004. *Capturing Campaign Dynamics: The Annenberg National Election Survey.* New York: Oxford University Press.

Ruppert, David, M. P. Wand, and R. J. Carroll. 2003. *Semiparametric Regression.* Cambridge: Cambridge University Press.

Shaw, Daron R. 1999. "A Study of Presidential Campaign Effects from 1952 to 1992." *Journal of Politics* 61:387–422.

Silverman, B. W. 1985. "Some Aspects of the Spline Smoothing Approach to Non-

Parametric Regression Curve Fitting." *Journal of the Royal Statistical Society,* Series B (Methodological), 47:1–52.

Verbeek, M., and T. Nijman. 1992a. "Can Cohort Data Be Treated as Genuine Panel Data?" *Empirical Economics* 17:9–23.

———. 1992b. "Pseudo Panel Data." In *The Econometrics of Panel Data: Handbook of Theory and Applications,* ed. Laszlo Matyas and Patrick Sevestre. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Wlezien, Christopher, and Robert S. Erikson. 2002. "The Timeline of Presidential Election Campaigns." *Journal of Politics* 64:969–93.

Zaller, John. 1998. "Monica Lewinsky's Contribution to Political Science." *PS: Political Science and Politics* 31:182–89.

———. 2002. "The Statistical Power of Election Studies to Detect Media Exposure Effects in Political Campaigns." *Electoral Studies* 21: 297–329.