

CAMPAIGN TRIAL HEATS AS ELECTION FORECASTS: EVIDENCE FROM THE 2004 AND 2006 CANADIAN FEDERAL ELECTIONS

Mark Pickup
University of Oxford
&
University of Calgary
mark.pickup@politics.ox.ac.uk

Richard Johnston
University of British Columbia
rjohnston@politics.ubc.ca

June 2006

An earlier version of this paper was presented to the 2006 Midwest Political Science Association Annual Meeting, Palmer House, Chicago, IL, 20-23 April 2006. We are grateful for astute comments and advice from Christopher Achen at the MPSA session. We are grateful to Daniel Rubenson who organised the presentation of this paper to a joint McGill University/ Université de Montréal meeting where very helpful suggestions were made. We are also grateful to our respective universities for research support. All responsibility for errors of fact or interpretation remains with the authors.

CAMPAIGN TRIAL HEATS AS ELECTION FORECASTS: EVIDENCE FROM THE 2004 AND 2006 CANADIAN FEDERAL ELECTIONS

ABSTRACT

The 2004 and 2006 Canadian elections saw unprecedented numbers of trial heat polls, including the first daily tracking polls in the country's history. This paper attempts to assess the competing claims regarding the accuracy of these polls by setting them in proper context and providing a systematic accounting of what might reasonably be expected from published results. We propose a method for disentangling sampling error, "house" effects, and true change. Most critically we estimate a value for bias in the polling industry as a whole.

Strikingly, systematic variation across firms within the industry is less notable than bias in the industry as a whole. There was less bias in 2004 than in 2006, but only marginally so. In each election, the industry underestimated the Liberal share and overestimated the NDP one. We speculate on sources of bias, and then ask if the discrepancy between polls and the electorate is methodological at the core or the product of last-minute strategic decision-making.

CAMPAIGN TRIAL HEATS AS ELECTION FORECASTS

The 2004 and 2006 Canadian elections saw unprecedented numbers of trial heat polls, including the first daily tracking polls in the country's history.¹ For each election, the predictive quality of the polls is controversial. The 2004 pattern is widely regarded as a poor showing, and the 2006 polling performance is in dispute. According to a *Montréal Gazette* headline (24 January 2006), "Pollsters missed their mark again: Liberal support again underestimated as most firms misread Canadian voters." In contrast, the *Toronto Globe and Mail* claimed on the same day that "pollsters' projections were close to the mark this time around." The *Globe* was particularly struck that, "unlike 2004, most major firms kept track of voters' intentions until the last minute."

This paper attempts to assess these competing claims by setting them in proper context and providing a systematic accounting of what might reasonably be expected from published results. We propose a method for disentangling sampling error, "house" effects, and true change. Most critically we estimate a value for bias in the polling industry as a whole.

Strikingly, systematic variation across firms within the industry is less notable than bias in the industry as a whole. There was less bias in 2006 than in 2004, but only marginally so. In each election, the industry underestimated the Liberal share and overestimated the NDP one. We speculate on sources of bias, and then ask if the discrepancy between polls and the electorate is methodological at the core or the product of last-minute strategic decision-making.

¹ The first public trackings, that is. Private firms have supplied forecasts and advice to the investment community for some years. Parties have deployed forms of tracking for at least 25 years. And the Canadian Election Studies have used a rolling cross section design (Johnston and Brady 2002) since 1988.

WHAT IS AT STAKE?

What is the value of evaluating the quality of election polls? Democracy, as it is commonly understood, requires informed voters.² Given their high accessibility, political polls are a potentially important source of political information. After the 2006 Canadian national election, Keith Newman, vice-president, public affairs at Environics Research Group Ltd defended the role of polls in election campaigns by arguing that they help voters vote with their “eyes wide open” (*Ottawa Citizen*, 1 February 2006). This, of course, can only hold to the extent that the information provided by polls is accurate. It is, therefore, important to consider the quality of this information, as we propose to do.

In addition to their potential value to democracy, polls are used by academics and policy makers alike to measure public opinion on a regular basis. We understand the statistical properties of random sampling error but we have no real sense of the potential bias that exists in such polling data. Rarely, do we have an opportunity to compare the information produced by these polls to real-world events. Election polls designed to predict a measurable outcome – the election result – give us a rare opportunity to do just this.

DATA

The data used in this paper are from the national election polls published in the major Canadian newspapers during the 2004 and 2006 election campaigns. The data were collected from newspaper accounts, from web sites of individual polling firms, and from web sites dedicated to the collection of this same information.³ For the 2004 campaign, 26 individual polls were identified. For 2006 the corresponding number was 66. The combined number of days in the

² For an interesting analysis of this view, see: Bernard Grofman (1995).

³ For example, <http://www.nodice.ca/>.

field for all polls was 80 in 2004 and 211 in 2006. Sample sizes varied greatly, ranging from 600 to 5254 in 2004 and from 400 to 2500 in 2006. In 2004, the firms contributing polls to our dataset were Ipsos Reid, Environics, Ekos, Compas, Leger, SES, and Pollara. In 2006, with the exception of Compas and the addition of Decima and Strategic Counsel, the same firms contributed polls.⁴

A FIRST COMPARISON

Figure 1 highlights differences between each year by presenting raw poll readings and a simple “poll of polls” for each party that won seats. “Raw” poll readings are in fact imputations: the same entry appears for all days a poll was in the field (hence the artificial identity of readings for consecutive days). The “poll of polls” is the average reading for all polls in the field on a given day weighted by sample size (which for any published poll is its total size divided by the number of days in the field).⁵

The first thing to catch the eye is the sheer density of readings for 2006, as compared to 2004. In the later year, a high frequency of readings appears right from the start and extends almost to election day; it helps that the 2006 campaign lasted almost two months. In 2004, one can speak of close coverage only for the last month of the writ period; the first week or so of the campaign is effectively missing. And 2004 polling ceased three days before election day (June 28th).

⁴ The 2004 and 2006 datasets can be obtained by contacting the authors.

⁵ We know that fieldwork is unlikely to be rectangularly distributed. We expect that the mode for completions will be the first day with a fairly rapid drop thereafter. See, for instance Johnston and Brady 2002. But in the absence of detailed reporting by commercial firms we have no alternative but to distribute observations rectangularly. An alternative imputation would have been to place a single observation at the midpoint of fieldwork. This makes little difference to the results. Some observations are for tracking polls, which publish overlapping information. Our imputations do not double count. Where possible, we use reported values for individual days. Otherwise, we report (and disaggregate) every n -th poll where n is the number of days pooled.

Striking for both years are discrepancies at the very end. These are signalled by the large circles on the rightmost side of each graph, which show parties' vote shares from official returns. Each year, the outstanding discrepancy is for the Liberal party: in actual votes, it was underpredicted by about three percentage points. The NDP was overpredicted each time. In 2004, the Bloc did better than predicted and the Conservatives worse. In 2006, each of these two parties did roughly as predicted by the very last polls.

In a sense, both journalistic claims cited at the outset for 2006 seem valid: 2006 was better for some parties and not worse for others. In particular, 2006 was better at predicting the share for the ultimate winner, the Conservative party. But it was hardly better for the Liberals and the NDP.

SOURCES OF MOVEMENT

Figure 1 certainly suggests that real movement occurred during the campaign, especially for the two parties realistically in competition to form the government. In each year, the Liberals started with the lead and seemed to lose it with three or four weeks to go. In 2004, they recovered, in a manner of speaking. But there is also ebb and flow that seems unlikely to reflect real forces. Some of this may reflect sampling error. But some may reflect practices specific to polling firms, “house effects” in the usual parlance. These can be a compound of sample design, weighting procedures, screening procedures (in search of the supposed “likely voter”), question order, question wording, and treatment of the “undecided”.

During and even more so after the 2006 election, there was a great deal of discussion about the potential effects of these types of differences among polling houses. Much of this discussion centred around the two firms that were in the field most often – SES and Strategic Counsel. SES consistently produced higher vote predictions for the Liberals than did Strategic

Counsel. To a lesser extent, it also tended to produce lower predictions for the Conservatives. Two key methodological differences have been identified as potential causes of the discrepancy. These methodological differences relate to question ordering and question wording (Vongdouangchanh and O'Malley, 2006). SES reported that it asked the vote intention question cold – that is, before making any other politically related inquiries. The concern is to avoid framing bias by potentially leading prior questions. In particular, the Strategic Counsel has been criticised for preceding (not necessarily immediately) the vote intention question with either a question regarding which party had the most momentum or a question asking if it was time for a change of government. It is clear how such questions could potentially bias the subsequent vote intention inquiry. In defence of this approach though, asking a cold vote intention question may not produce a realistic response. The actual vote is likely made in the context of such political debates (momentum and time for a change). Warming the respondent up to the vote intention question may put her in a mindset that is more reflective of Election Day.

The second identified methodological difference between SES and Strategic Counsel is that the former asked an open-ended vote intention question while the Strategic Counsel named each party by name.⁶ Asking an open-ended question (not prompting the respondent with the names of the parties) can be effective at filtering out those that are unlikely to vote. A respondent that cannot name a party a month or a week before the election is less likely to actually go out and vote than a respondent which can. With a closed ended question, if such a respondent felt uneasy admitting their lack of interest or intention to vote and were presented with a list of parties, they may be tempted to simply name one on the list. They cannot do this with an open-ended question. When presented an open-ended question, however, a respondent who is unlikely

⁶ Examples of the question wording used by various firms are included in Web Appendix 1.

to vote but has some minimal political knowledge may simply respond with the first party that comes to mind. This is likely to be the party of the government and to a lesser extent the party of the opposition. This would bias vote predictions in favour of the government and potentially other major parties.

Other major methodological differences can be identified between firms other than the SES and Strategic Counsel. Unlike any other major firm, Ekos asked respondents which party they intended to vote for in “the January 23 election.”⁷ All other firms asked respondents which party they would vote for or preferred if an election were held that day or the following. In terms of the screening procedures, Pollara differed from other firms by only including those respondents which indicated that they were very likely to vote in their vote intention calculation.⁸ Leger was at the opposite extreme, in calculating vote intention shares they redistributed the undecideds proportionally rather than simply leaving them out of the calculation.⁹ A middle strategy used by a number of other firms was to ask undecided voters a follow-up question regarding which party they were leaning towards and included such “leaners” in their vote intention calculations but excluding those that remained undecided after the second question. Leger was also distinct in terms of weighting on first language spoken at home, unlike other firms that generally all weighted on demographics such as age, gender and region.¹⁰

Similar differences to all of these identified in 2006 existed in the 2004 election campaign polls. In principle, it would be best to treat each dimension of house practice as a variable. But descriptions of house practice are murky, and in the analysis described below the best we can do is treat the methods of each house as a dummy variable.

⁷ Ekos website <http://www.ekos.com/main.asp>, accessed May 1, 2006.

⁸ Pollara website <<http://www.pollara.ca/>>, accessed December 11, 2005.

⁹ Leger website <<http://www.legermarketing.com/eng/>>, accessed December 8, 2005.

¹⁰ At the very least, the other polling firms made no indication in their press releases or on their web sites that they weighted by language. This, of course, does not guarantee that they didn't.

Discrepancies at the end between trial heats and the actual outcome also raise the question of bias in the industry as a whole. There is certainly a suggestion of such bias each year. The difference between the final poll of poll prediction and the actual election outcome is statistically different for the NDP and the Liberals in both 2004 and 2006, and the Conservatives and Bloc Québécois in 2004.¹¹ If it seems unlikely that the industry as such is biased because of individual polling house methodical choices, it could be that voters interact with polls in ways that are misleading about the likely final result. Strategic considerations may not be reflected in the polls. Survey samples could misrepresent turnout likelihoods. Or increasingly low response rates to commercial polls may produce a misrepresentation of the likely voting public – e.g., a biased sample. It is important to keep in mind that while these sources of industry bias may not be the consequence of the particular methodical choices of any individual firm, they are still methodological problems. And their solutions are likely methodological. A source of industry bias that is less about methodology is the potential problem of a large portion of the voting public making their vote decision in the polling booth. If the distribution of votes for these individuals differs greatly from the larger voting public, for strategic reasons or otherwise, this could produce a marked difference between poll predictions and actual outcomes.

Before we go looking for industry-wide explanations, however, we need to dismiss simpler possibilities telegraphed in earlier paragraphs. It could be that the discrepancy between late polls and election day is within the bounds of sampling error, as indicated by the usual tests and criteria. Or it could be that late polls are dominated by firms that have a systematic house-

¹¹ The 95% confidence interval for the final poll of poll prediction for the Liberals in 2006 is 25.5-28.2%. The actual election outcome was 30.1%. In 2004, the Liberal confidence interval is 31.9-34.5% and the outcome was 36.7%. The corresponding final poll of poll prediction confidence intervals and final vote shares for the NDP are 18.5-20.9% and 17.5% in 2006, and 18.1-20.3% and 15.7% in 2004. In 2004, the final poll of poll prediction confidence interval for the Conservatives is 30.4-32.9% and the final vote share was 29.6%. For the Bloc Québécois, the final poll of poll prediction confidence interval is 10.3-12.1% in the final vote share was 12.4%.

specific bias. It is necessary, therefore, to utilise a more sophisticated method of determining industry bias. In particular, one that also estimates individual house biases and does so while accounting for sampling error.

To account for any single component of survey dynamics, we need to account for all. To that end, we propose to extract house bias and sampling error simultaneously, by modeling observed public opinion as a composite of three unobserved components. The model is in state-space form and is estimated by Bayesian analysis. An example of such an approach is Jackman (2004), although our approach is not necessarily the same as his.¹² We deploy the state-space model as follows:

$$\begin{aligned} share_t &= \alpha_t + \delta_t + v_t \\ \alpha_t &= \rho\alpha_{t-1} + \varepsilon_t^\alpha \\ \delta_t &= \sum_i \beta_i P_{it} \end{aligned}$$

where *share* is a party's measured share of decided vote intention on day *t*;

α_t is the party's true deviation from its equilibrium vote share at *t*;

v_t is the sampling error component at *t*;

δ_t is the equilibrium vote share including bias from the firms in the field at *t*;

β_i is the bias for the *i*th firm; and

P_{it} is the *i*th firm's share of the total sample in the field at *t*.

ρ is the first-order autoregressive component (about which more below); and

ε_t^α subsumes all unmeasured variation operating at *t*.

¹² House and sampling error effects can also be extracted sequentially, on the example of Erikson and Wlezien (1999). In this strategy, house effects are estimated in an OLS setup with each day also a dummy variable (with no intercept term). From coefficients for individual days and the coefficient for the median firm a house-controlled daily series can be derived. This series can then be smoothed, by filtering or by state-space estimation as the analyst prefers. We argue (Pickup and Johnston 2005) that this may overestimate house effects. As it happens, for the Canadian cases each strategy yields results broadly similar to those reported here.

The variance of the sampling-error component each day can be estimated from the size of that day's sample.¹³ Also, to the extent that public opinion contains memory, readings from surrounding days can be used to increase the accuracy of each day's measurement. This allows us to control for the noise produced by this error and to estimate the real public opinion component.¹⁴ The real public opinion component is the value of α_t plus an equilibrium value. Explicating the meaning of this equilibrium first requires discussion of δ_t and $\sum_i \beta_i P_{it}$. The goal of controlling for house bias is to set the δ_t term equal to the true equilibrium value of vote share, excluding house bias. To do this we require estimates of β_i for each i , where β_i is the bias for house i relative to real public opinion. Before the election, of course, we do not have a benchmark for real opinion. But we can stipulate a benchmark firm, estimate the deviations of other houses from the benchmark, and then normalize the series to the median house. This would provide the median house equilibrium which can be added to the α_t series in order to calculate public opinion as it would be predicted by this 'average' firm, controlling for sampling error.

This discussion underlines what we can only assume: that the industry as a whole converges on the truth. After the fact we do know the truth, in a manner of speaking: the votes actually cast on Election Day. The election result can be added to the series, in effect as a poll with zero measurement error and zero house bias. This provides estimates of each house's bias

¹³ The standard deviation of the estimated sampling error is calculated as $\sqrt{p_t(1-p_t)/N_t}$, where p_t is the proportion of valid respondents indicating a vote intention for the relevant party at time t and N_t is the sample size based on the number of decided voters polled.

¹⁴ Estimation used the program *Winbugs*, which uses Markov Chain Monte Carlo (MCMC) methods. In our estimations a single chain with 60,000 iterations was used. For each parameter, an uninformative prior distribution was used. It is an interesting question as to whether a Winbugs analysis that uses uninformative priors is really Bayesian. A fairly compelling argument can be made that it is not and that the result is no different than that from a maximum-likelihood estimation. In this particular context, the Bayesian aspect of the analysis is found in the daily updating of vote share predictions with new polling data using previous polling data as a prior. In this sense, although a "flat prior" is used for public opinion at $t = 0$, a "proper" prior is used for calculations at all subsequent time points.

relative to this ‘perfect poll’ of public opinion. A new predicted series can be derived from this anchoring of public opinion to the election result. This is our best estimate of actual public opinion throughout the campaign. Actual public opinion is conceptualised as that produced by combining the equilibrium value of the hypothetical polling firm that produced the zero error, zero bias election result with the α_t series. Visual or arithmetic comparison of this estimate of actual public opinion and the estimated median house series gives an estimate of industry bias. Arithmetically, it is simply the difference in the estimated equilibrium values.

Also requiring comment is the presence of a first-order autoregressive parameter. The nature of the link, if any, between successive observations is highly consequential. Jackman (2004), for example, posits as a simplifying assumption that his Australian opinion series is fully integrated. If this assumption corresponds to the facts, then the time series has perfect memory: the effect of any shock is entirely preserved; only another shock will displace the system. In fact, this assumption is dispensable, and the degree of integration can be estimated directly. We can then treat the degree of integration as a matter for direct estimation rather than *a priori* assumption.¹⁵

FINDINGS

Table 1 presents the results that pertain to house and industry bias. To get a sense of differences among houses, consider the estimations labeled “median.” For the most part, firms cluster quite closely around the median. In 2004, three firms differed significantly from the median, however in one instance only for Conservative intentions, in another only for NDP intentions and in the

¹⁵ As it happens, values estimated for ρ in our estimations lie close to one (0.98) for all parties in 2006. In 2004, there is not enough movement left in the NDP, Conservative and Bloc Québécois timeseries, once house effects and sampling error have been controlled, in order to produce a precise estimate of ρ . For further discussion on this and the general dynamics of the 2004 campaign see Web Appendix 2.

third only for Bloc intentions. Similarly in 2006, only three instances of significant house effects appear. Strikingly SES is responsible for two of these. The SES significantly predicted higher vote shares for the Liberals than did the industry median and lower vote shares for the NDP. In each case, the industry median did a poor job of forecasting the party's result. Although one firm stood out from the rest each election, there was no continuity from 2004 to 2006 in that firm's identity.

If house effects are modest, industry bias is striking. This is shown by the columns labeled "actual," which are the products of estimation fixed to election day as the last observation. The difference term at the bottom of the table is, in effect, the coefficient for the median house (or houses). In each year, the Liberal share is underestimated by at least three points and the NDP share, overestimated by at least two points. Industry bias seems smaller in 2006, but only by a fraction. Figure 2 makes the point visually, with daily imputations and 95% confidence bands for the median estimations. It shows that for each year the actual result is outside the confidence interval for both the Liberals and the NDP. No such problem occurs for the Conservatives and Bloc. The disappearance of the 2004 industry biases in favour of the Conservatives and against the Bloc Québécois suggested by their respective poll of polls, demonstrates the value of our procedure. The results suggested by each poll of polls are subject to sampling error and affected by which polling house happens to be in the field near the end of the election campaign.

DISCUSSION

It is natural to wonder if observers who saw poll predictions in 2006 as better than 2004 were focusing on predictions of the winner. By the end of the 2004 campaign, the Liberals were seen

as ahead by most polls,¹⁶ but the Conservative and Liberal imputations had a large amount of overlapping error. It is easy to see why the race seemed too close to call. In the end, of course, it was not close. The opposite problem bedeviled trial heats in 2006, but in doing so it made the result seem clear. There seemed to be no chance that the Conservatives would lose, at least in the sense of the popular vote plurality. But the underestimation of the Liberals tempted some to forecast an outright seat majority for the Conservatives.¹⁷ It is also tempting to suggest that in 2006, SES saved the industry's bacon. It regularly suggested greater Liberal strength than most other firms did, and it generated more independent observations than others.

Regardless of the magnitude of industry bias in 2004 and 2006, the fact that bias was significant and consistent raises the question of whether it was due to the methodological choices of individual firms or was the product of unobserved or even unmeasurable voter considerations.¹⁸ On the side of methodological errors, it is possible that the choice by many polling firms to frame the vote intention question biased the estimated vote share for the incumbent Liberal government downward. This may have occurred by priming the respondents to think in terms of a need for governmental change, tapping into underlying feelings that the Liberals had become arrogant.¹⁹ That this bias against the Liberals translated into an overestimation of NDP support may simply be because that in the process of pushing Liberal voters away from indicating their true vote intention, this particular methodological choice pushed them to indicate their second choice (the NDP) as their vote intention.

¹⁶ At least by the most *observations*, not necessarily by most firms.

¹⁷ See the predictions posted at <http://www.trendlines.ca/politics.htm#2004>.

¹⁸ Some voter considerations may be inherently unmeasurable, while others may simply not be observed by current methodological approaches. In this sense, the distinction between the two sources of industry bias is rather blurred. If the perfect methodological research design is realisable then all sources of error are methodological. Here we make a distinction between bias that firms can realistically be expected to be able to control and bias produced by factors that we feel are currently beyond their control.

¹⁹ See, for instance, the discussion in Vongdouangchanh and O'Malley (2006), p. 70.

Industry bias from methodological errors could also reflect poor filtering of those that ultimately did not vote.²⁰ By including such respondents in the vote intention calculation, the segment of the population which feels disenfranchised may have been over sampled. They would be over sampled because they would be disinclined to actually vote and therefore should not be counted as strongly in a calculation of actual vote share. This over sampling would produce a bias against the Liberals as this population would also be disinclined to indicate the incumbent government as their vote choice.

Industry bias may have been produced by considerations outside firms' purview. A fair bit of discussion in the post-2004 election media coverage suggested that a number of voters (particularly in Ontario) decided to vote Liberal at the last moment, possibly for strategic reasons.²¹ These individuals might otherwise have voted NDP. Last-minute strategic decision-making would be difficult to catch, either because the respondents in question switched parties outright after the last poll was conducted or because they shifted from the undecided category disproportionately to a Liberal vote.²² This would explain the underestimation of Liberal vote share and the overestimation of NDP vote share. In 2004, polling firms discontinued fieldwork before the last weekend and so ruled themselves out of catching any late shifts.²³

In 2006, in response to the difficulties in 2004, some firms polled right up to Election Day. As it turns out, the resulting estimate of the 2006 industry median prediction, controlling for individual house effects and sampling error (figure 2), provides little evidence of any last-

²⁰ This is essentially a disproportionate turnout problem. There is a correlation between the likelihood of showing up to vote and supporting the incumbent Liberals – a positive correlation.

²¹ Vongdouangchanh and O'Malley (2006), p. 70 and Ipsos-Reid (2004), pp. 5-6.

²² The percentage of undecided voters remained high in the 2004 polls right up to the last polls conducted, often around 20%.

²³ The CES team polled a day longer and there is some indication of last-minute shifting to the Liberals (see Web Appendix 2). However, the small day-to-day sample sizes of the CES puts this shift within a 95% margin of error, even after borrowing strength from temporally close polls.

minute shifting of vote share. For last-minute decision making to be the cause of industry bias, the four to three percentage point gap would all have to be the product of decisions made in the actual polling booth on election day itself. If this is the case, polling firms have a very real problem on their hands. To the extent that last-minute strategic voting with a dominant direction occurs, trial heat polls may never predict outcomes accurately. Post-election analyses suggest that nine to twelve percent of voters made their decision in the polling booth.²⁴ Allowing for roughly 10 percent of the population to have made its vote decision on election day, sixty percent of those that did so would have had to have voted Liberal in order to close the gap between the industry median prediction and the actual election outcome. This is twice the occurrence of Liberal voting than occurred in the general population. It would seem that last-minute decision making is unlikely to be the only source of bias. So why else could have industry bias persisted in 2006?

The industry's claim in 2004 was one of last-minute strategic voting. While, closer inspection suggests that it is unlikely to have been last-minute voting that caused the problems in 2006, this does not rule out the issue of strategic voting. Not, at least, if we allow for the possibility that the trial heat polls are unable to pick up on strategic voting even when it is not a product of last-minute decision making. While still a major problem for the polling industry, this is a source of bias that should have a methodological solution – at least in theory. However, in contradiction to the possibility that strategic voting was the source of bias in 2006 is the fact that SES had one of the smallest biases against the Liberals or in favour of the NDP. SES asked respondents what their party preference was, rather than vote intention as other polling firms did.

²⁴ An Ipsos-Reid online exit poll produced the estimate of 9% (*Vancouver Sun* "This Election Was Decided by Christmas," January 24, 2006) while a SES telephone poll produced an estimate of 12% (see Nanos, 2006).

Such question wording should exacerbate any bias produced by the presence of strategic voting, not reduce it.

It is not immediately clear how to determine if industry bias in 2004 and 2006 was the product of methodological choices of individual polling firms or of strategic considerations in voters' minds. The next Canadian federal election, with the Conservative party as the incumbent government, could be the test case, however. If underestimation of Liberal support in 2004 and 2006 was the product of methodological choices that biased responses against the incumbent government, we should see a new bias against the incumbent Conservatives emerge in the next election. If underestimation of Liberal support was a consequence of firms being unable to observe last-minute strategic decisions or strategic decisions in general, then the bias against the Liberals and toward the NDP should hold even as the forecast for Conservative intentions remains sound. Unfortunately, these two possibilities do not exhaust the potential sources of bias. If other yet to be defined sources of bias are at work, the next Federal election may not provide the elucidation that we might hope.

Whatever the process, there does seem to be a case to be made for polling to the bitter end. The more firms that do so, the less subject forecasting will be to the choices of one or a small number of firms. The larger the number of firms in the field, the more offsetting there should be of house effects and the larger the number of observations that will be available to pool.²⁵ Polling till the end may account for the one percentage point decline in industry bias against the Liberals between 2004 and 2006. Unfortunately, additional firms polling right up until election day may or may not remove industry bias. If this bias is a product of last-minute

²⁵ Gains from polling down to the end will be maximized if each firm clears its samples consistently from day to day, such that the last day's sample take is as much of a random draw as should be the case for preceding days. For a forceful argument along these lines, see Johnston and Brady (2002).

strategic considerations, polling to the end can only capture the decisions made before election day itself and polls will continue to be subject to some bias. Even then, polls, as they are currently conducted, may be unable to capture strategic considerations even when they are not last-minute. This is a methodological problem that polling firms need to consider. Moreover, if industry bias is a product of methodological choices, such as question wording, polling to the end will not correct the problem and other solutions ought to be sought out.

Figure 1. Daily Readings and “Poll of Polls”

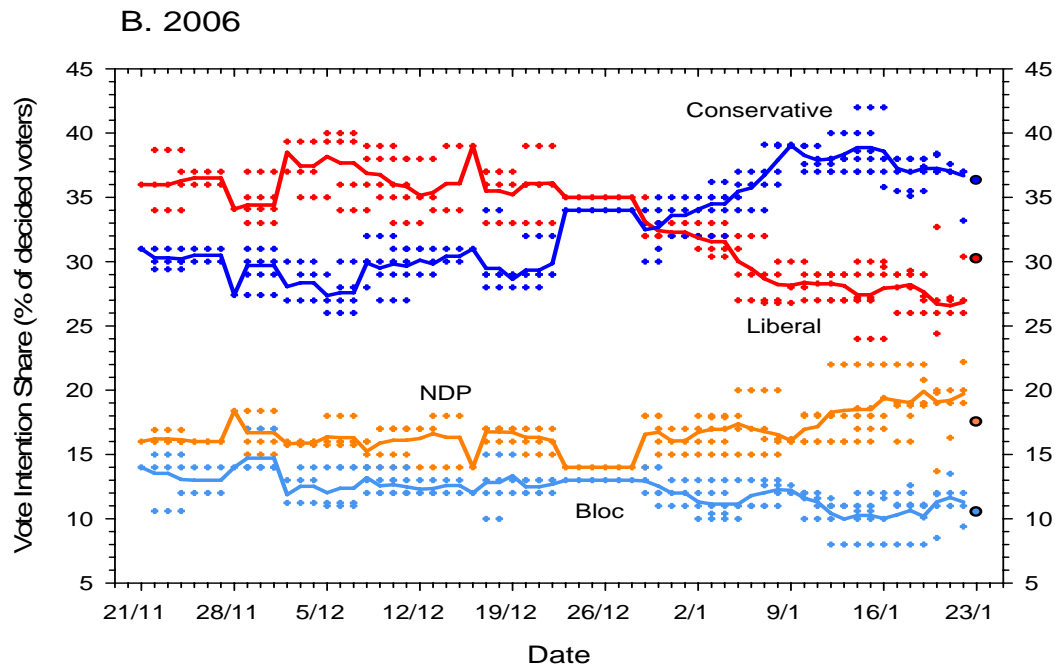
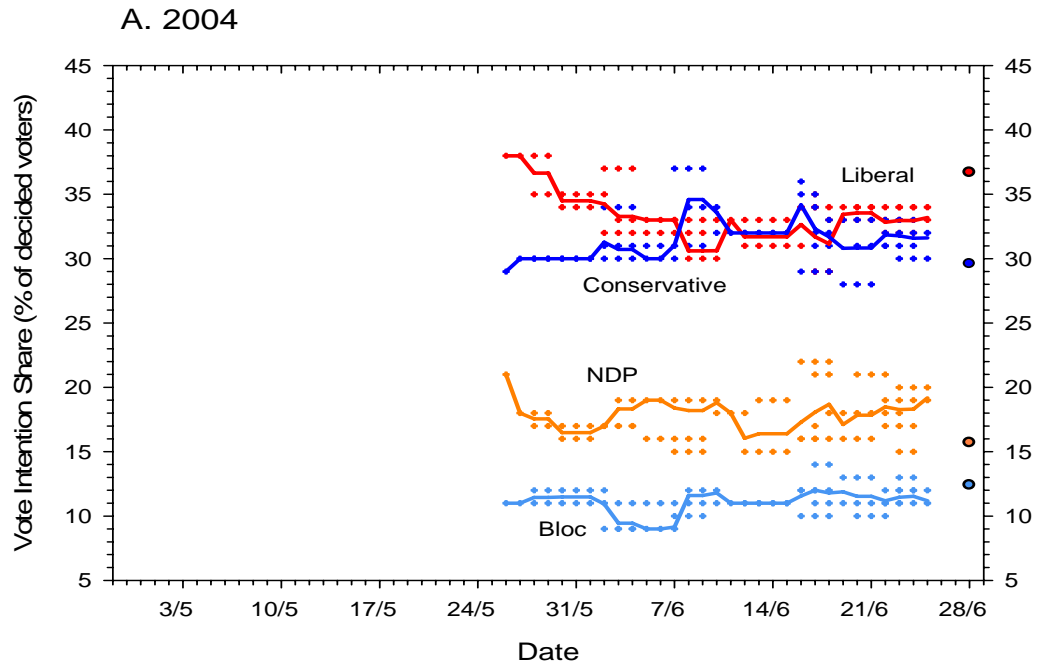


Figure 2. Smoothed Plots (Industry Median) and 95% Confidence Intervals

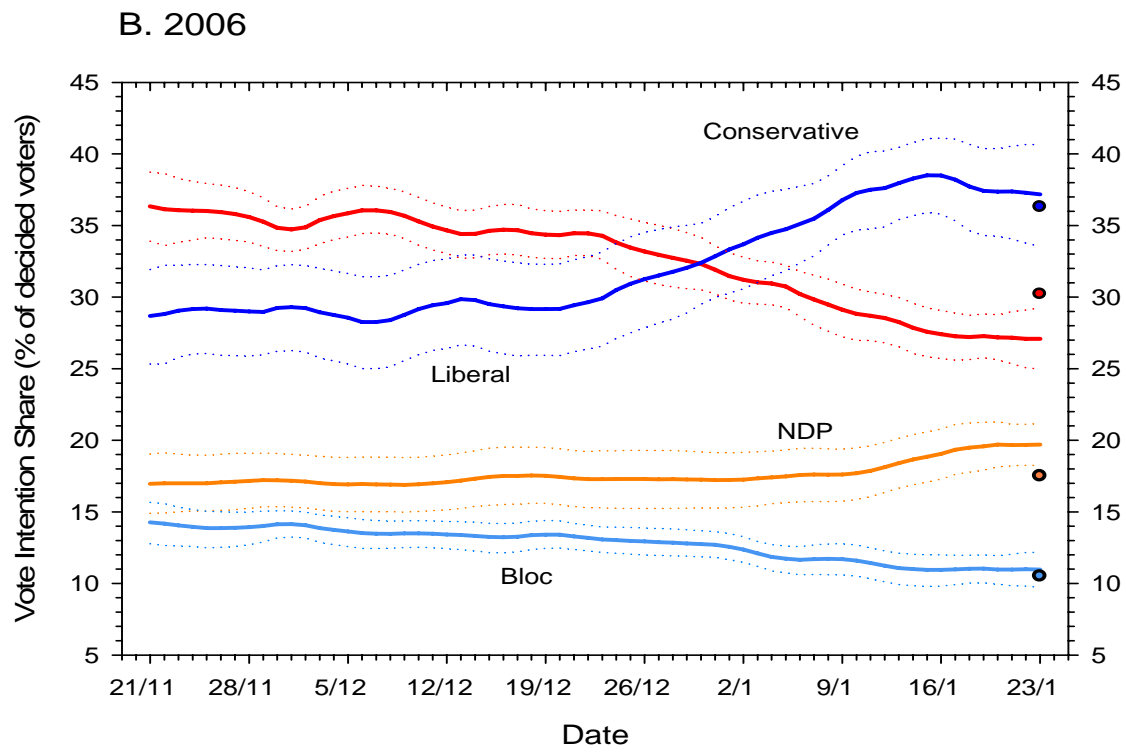
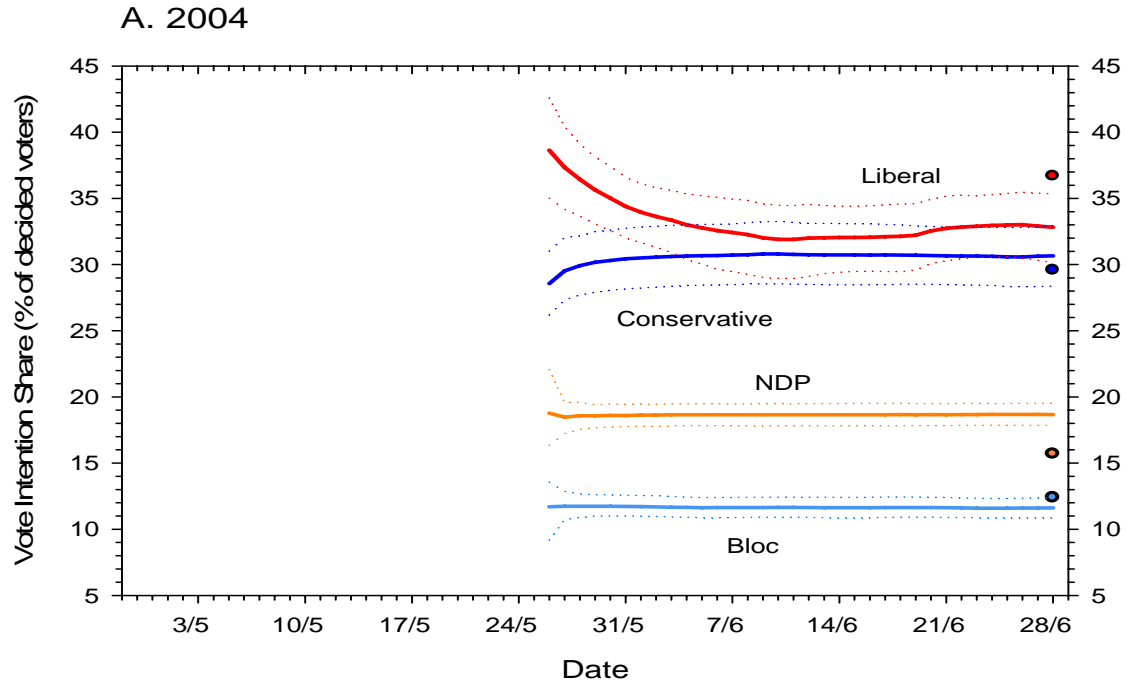


Table 1: Estimated Polling House and Industry Bias, 2004 & 2006

Polling House	Liberals				Conservatives				NDP				Bloc			
	Bias Relative to:				Bias Relative to:				Bias Relative to:				Bias Relative to:			
	2004 Median	2004 Actual	2006 Median	2006 Actual	2004 Median	2004 Actual	2006 Median	2006 Actual	2004 Median	2004 Actual	2006 Median	2006 Actual	2004 Median	2004 Actual	2006 Median	2006 Actual
SES	-0.2 (2.63)	-4.2** (1.91)	2.1** (1.02)	-0.8 (1.22)	3.5 (2.01)	0.9 (1.63)	2.5 (1.63)	3.4** (1.33)	0.2 (1.40)	3.2** (1.18)	-3.3** (1.03)	-1.1 (0.80)	-1.8 (1.18)	-2.6** (1.00)	0.0 (0.39)	0.5 (0.71)
Strategic Counsel	—	—	-0.3 (0.67)	-3.3 (1.22)	—	—	0.5 (1.67)	1.5 (1.34)	—	—	-0.3 (1.20)	1.9** (0.81)	—	—	-0.3 (0.72)	0.2 (0.70)
Ekos	-0.5 (1.18)	-4.4** (0.81)	-2.2 (1.38)	-5.2** (1.48)	1.7 (1.22)	2.7** (0.59)	0.4 (1.17)	1.3 (1.62)	0.0 (0.00)	3.0** (0.42)	-0.1 (0.82)	2.1** (1.01)	0.0 (0.00)	-0.8** (0.37)	0.5 (0.92)	1.0 (0.86)
Leger	0.6 (1.45)	-3.3** (0.98)	2.0 (1.06)	-0.9 (1.19)	-0.6 (1.16)	0.4 (0.59)	-2.2 (1.62)	-1.3 (1.31)	0.1 (0.55)	2.0** (0.42)	0.4 (0.91)	2.6** (0.78)	-2.5** (0.45)	-3.3** (0.37)	-2.5** (0.66)	-2.0** (0.68)
Pollara	0.9 (2.67)	-3.0 (2.74)	0.7 (1.54)	-2.3 (1.69)	6.4** (2.67)	7.4** (2.43)	2.9 (1.88)	3.8** (1.84)	-3.2 (1.89)	-0.2 (1.90)	-1.7 (1.21)	0.5 (1.13)	2.5 (1.62)	1.7 (1.62)	-1.1 (0.94)	-0.7 (1.02)
Decima	—	—	-3.3 (2.24)	-6.3** (2.31)	—	—	-0.4 (1.17)	0.6 (2.54)	—	—	1.4 (1.64)	3.6** (1.61)	—	—	0.1 (1.39)	0.6 (1.39)
Ipsos Reid	0.0 (0.00)	-3.9** (1.23)	0.8 (0.67)	-2.7** (1.27)	-0.1 (1.69)	0.9 (0.95)	-0.4 (1.17)	0.3 (1.40)	-3.7** (0.75)	-0.8 (0.73)	0.3 (1.11)	2.5** (0.87)	0.5 (0.64)	-0.3 (0.62)	0.0 (0.39)	0.5 (0.74)
Environics	0.3 (1.96)	-3.6** (1.43)	-0.9 (2.17)	-3.9** (1.60)	0.0 (0.00)	1.1 (1.15)	-1.9 (2.66)	-0.9 (1.76)	1.6 (1.03)	4.6** (0.93)	0.1 (0.82)	2.3 (1.20)	-0.3 (0.85)	-1.1 (0.79)	1.3 (1.22)	1.7 (0.99)
Compas	-0.4 (5.29)	-4.2 (5.58)	—	—	1.6 (4.69)	2.7 (4.65)	—	—	3.7 (3.87)	6.7 (3.70)	—	—	3.2 (3.32)	2.4 (3.19)	—	—
Industry Bias		-3.9** (1.23)		-3.0** (1.06)		1.1 (1.15)		1.0 (1.78)		3.0** (0.42)		2.2** (0.74)		-0.8** (0.37)		0.5 (0.62)

- Notes: 1) values in round parentheses are standard deviations;
2) ** p < 0.05;
3) industry bias is calculated as the unweighted median polling house bias

REFERENCES

- Erikson, Robert S. and Christopher Wlezien. 1999. "Presidential Polls as a Timeseries: The Case of 1996." *Public Opinion Quarterly* 63: 163-177.
- Grofman , Bernard N. 1995 *An Economic Information, Participation, and Choice: An Economic Theory of Democracy in Perspective*. Michigan: University of Michigan Press.
- Ipsos-Reid. 2004. "For the Record: 2004 Federal Election Polling Post Mortem." Toronto: online manuscript (accessible at www.ipsos.ca)
- Jackman, Simon. 2004. "Polling and Smoothing the Polls over an Election Campaign." Stanford University: unpublished manuscript.
- Johnston, Richard, and Henry E. Brady. 2002. "The Rolling Cross Section Design." *Electoral Studies* 21: 283-95.
- Nanos, Nikita James. 2006. "SES Nation State Poll – Federal Election 2006 – What Happened?" SES e-mail update, February 23.
- Pickup, Mark, and Richard Johnston. 2005. "Measurement Error and House Bias in 2004 Presidential Campaign Polls." University of Oxford: unpublished manuscript.
- Vongdouangchanh, Bea and Kady O'Malley. 2006. "Inside the Poll Story: Who Got It Right, Who Got It Wrong, and Why?" *Policy Options* March: 68-73.

WEB APPENDIX 1: EXAMPLES OF VOTE INTENTION QUESTION WORDING

Pollara

1. If a federal election were held today, how likely would you be to vote? Would you be very likely, somewhat likely, not very likely, not at all likely?
2. (ONLY THOSE THAT ARE VERY LIKELY TO VOTE) If a federal election were held today, which party would you most likely vote for? (READ and ROTATE)
Liberal Party - Paul Martin
Conservative Party- Stephen Harper
NDP - Jack Layton
(Quebec Only) Bloc Québécois - Gilles Duceppe

SES

1. If a federal election were held today, could you please rank your top two current local voting preferences? (First ranked reported)?
2. Are you currently leaning towards any particular federal party, and if you are, which party would that be?

Strategic Counsel

Asked only of decided voters:

1. If the election was being held tomorrow, do you think you'd be supporting the (ROTATE LIST) Liberal candidate in your area, Conservative candidate in your area, the NDP candidate in your area, or the Green party candidate in your area or (QUEBEC ONLY) Bloc Québécois candidate in your area?
2. In that case, which party's candidate in your local area would you be leaning towards at this time? Would it be the Liberal candidate in your area, Conservative candidate in your area, the NDP candidate in your area, or the Green party candidate in your area or (QUEBEC ONLY) Bloc Québécois candidate in your area?

Ekos

Which political party do you intend to vote for in the January 23 election?

Leger

1. If FEDERAL elections were held today, for which of the following political parties would you be most likely to vote for. Would it be for ...?
READ AND ROTATE - ONE MENTION POSSIBLE
...Paul Martin's Liberal Party
...Gilles Duceppe's Bloc Québécois

...Stephan Harper's Conservative Party
...Jack Layton's NDP
...Jim Harris' Green Party
Other party
Would not vote/would spoil ballot
Don't know
Refusal
(IF UNDECIDED, ASK LEANING QUESTION)

2. Even if you have not yet made up your mind, which political party would you be most likely to vote for? Would it be for...

READ AND ROTATE - ONE MENTION POSSIBLE

...Paul Martin's Liberal Party
...Gilles Duceppe's Bloc Québécois
...Stephan Harper's Conservative Party
...Jack Layton's NDP
...Jim Harris' Green Party
Other party
Don't know
Refusal
(In calculating vote intention shares, undecideds were redistributed proportionally.)

WEB APPENDIX 2: ESTIMATED ρ AND CAMPAIGN DYNAMICS

Web table 1 contains the estimated AR(1) terms for each of the party's vote intention series in 2004 and 2006. In 2006, the estimated AR(1) term is 0.98 for each party. In 2004, the Liberal AR(1) term is smaller at 0.80 and the standard deviation of the distribution of this parameter is eight times larger than it was in 2006. More notable, are the small AR(1) terms estimated for the NDP, Conservatives and Bloc in 2004 and the corresponding large standard deviations. In fact, the standard deviations are so large that the AR(1) terms cannot be said to be statistically different from zero.

The source of imprecision in the estimates of the 2004 NDP, Conservative and Bloc AR(1) terms is indicated by the variances of the estimated α components. This reflects the movement in vote intention that remains once sampling error and polling house effects are removed. Except for the Liberals, the 2004 variances are 100 to 1000 times smaller than in 2006. Once sampling error and polling house effects are controlled, there is very little movement in vote intention for the NDP, Conservatives and the Bloc. This leaves very little data by which to estimate memory in public opinion. Put another way, the lack of change in public opinion makes it very difficult to describe the nature of public opinion dynamics.

The lack of movement in public opinion in 2004 is reflected in the relatively flat lines in figure 2a. The only substantial movement observable in this plot is for the Liberals at the beginning of the campaign. Typical of many Canadian federal election campaigns, the incumbent government lost some ground once the election campaign began and moved very little after. This reading of the 2004 election campaign belies the standard interpretation that it was an unusual event. In contradiction to reports in the media, the Conservatives never did surge ahead of the

Liberals and were never in a position to win the election let alone form a majority.²⁶ There is also no real evidence of the interpretation provided by many polling firms and academics alike that suggests the Liberals won the election through a last-minute recovery.²⁷ Figure 2a ends three days before election day but even examining the smoothed 2004 CES data (Web figure 1) which takes us within two days, the only shift in Liberal support observable at the end of the campaign is not statistically significant at a 95 % confidence level.²⁸ In fact, none of the movement in the CES tracking of Liberal government support is outside of the 95% confidence boundary for the beginning of the period.

²⁶ e.g., “Tories Could Be Edging Towards Majority,” *Vancouver Sun* (June 12, 2006).

²⁷ In particular, see: “How the Race Was Won,” E. Gidengil, A. Blais, J. Everitt, P. Fournier and N. Nevitte, *The Globe and Mail* (July 14, 2004). A useful discussion on this from the perspectives of both the polling firms and academics is provided by Barry J. Kay and Chris Cattle (2005) “Advanced Polls and the ‘Late Switch’ in the 2004 Election.” *Canadian Parliamentary Review*: 18-23.

²⁸ The purpose of smoothing is to control for movement produced by measurement error. This is important given the small day-to-day sample sizes of the CES tracking..

Web Table 1: AR(1) Values

	2004		2006	
	ρ	VAR(α)	ρ	VAR(α)
Liberal	0.80 (0.074)	2.684	0.98 (0.009)	10.770
NDP	0.14 (0.478)	0.001	0.98 (0.033)	0.777
Conservatives	0.58 (0.385)	0.183	0.98 (0.014)	13.863
Bloc	0.30 (0.491)	0.001	0.98 (0.013)	1.274

Web Figure 1: 2004 Liberal Vote Share

2004 Liberal Vote Share (Kalman Smoothed)
Source: 2004 CES

