J. R. Statist. Soc. A (1996)
159, *Part* 3, *pp.* 535–545

Public Opinion Polls: the UK General Election, 1992

By T. M. F. SMITH†

University of Southampton, UK

[Received September 1995. Revised February 1996]

SUMMARY

The public opinion polls conducted before the 1992 general election were a statistical disaster. The errors were far in excess of expected sampling variation and can only be explained by biases specific to the context of the 1992 election since previous election results had been forecast accurately. The replication in the polls in the months before the election provides a natural experiment from which the sampling variance of quota samples can be estimated. Using a components-of-variance model, company-specific variances are estimated and are found to be consistent with those which would have been obtained from equivalent random sample designs.

Keywords: BIAS; COMPONENTS OF VARIANCE; DESIGN EFFECTS; MULTILEVEL MODEL; QUOTA SAMPLES; QUOTA SAMPLING VARIANCE

1. INTRODUCTION

It is rarely possible to validate a statistical analysis based on observational data, such as a sample survey. One of the rare exceptions is public opinion polling before an election. Whatever the pollsters may say the objective of the polls is to provide relevant data on which to predict the outcome of the election. When experts use the data from the polls and make correct predictions then the pollsters claim success, and when they use the data from the polls and fail to make correct predictions then the polls may be partly to blame. When, as in the UK in 1992, all the polls are demonstrably wrong, then it is difficult for the pollsters not to accept responsibility. But before criticizing the pollsters too strongly statisticians should ask themselves whether any of their statistical exercises have been exposed to such a rigorous examination. Are they sure that their experiments and surveys would perform as well as the polls have done in the past?

Despite this *caveat* about rushing too hastily into judgments about the polls, the failure in the UK in 1992 really was quite spectacular. Over 50 polls were conducted in the month before the general election (see Table 1), and they gave leads to the Conservative Party (under John Major) over the Labour Party (under Neil Kinnock) of between -7 percentage points and $+5$ percentage points, with an average of about -1.5 percentage points. The actual result was a Conservative lead over Labour of $+7.6$ percentage points. A sign test suggests that either we have observed an event with probability 2^{-50}, a rather small number, or that the polls were biased. The interpretation is not in doubt and the presence of bias was immediately accepted by the Market Research Society (MRS) of which all the major polling companies are

†*Address for correspondence*: Faculty of Mathematical Studies, University of Southampton, Highfield, Southampton, SO17 1BJ, UK.
E-mail: tmfs@maths.soton.ac.uk

0035–9238/96/159535

TABLE 1
Published national opinion polls in the 1992 general election†

Fieldwork	*Company*‡	*Source*	*Sample*	*Conservative vote (%)*	*Labour vote (%)*	*Liberal Democrat vote (%)*	*Conservative lead (%)*
March 11th	NOP	*Mail on Sunday*	1050	41	40	15	1
March 11th–12th	MORI	*Times*	1054	38	41	16	−3
March 11th–12th	MORI	*Sunday Times* (P1)	1544	40	39	18	1
March 11th–13th	Harris	*Observer*	1054	40	43	12	−3
March 11th–13th	Harris	*Daily Express*	1086	39	40	16	−1
March 11th–13th	Harris	London Weekend Television	2186	37	41	17	−4
March 12th–13th	NOP	*Independent on Sunday* (P1)	2155	40	41	14	−1
March 13th	ICM	*Sunday Express*	1059	39	40	16	−1
March 15th–16th	Harris	*Daily Express*	1081	41	38	17	3
March 16th	MORI	*Times*	1099	38	43	16	−5
March 17th	ICM	*Guardian*	1100	38	43	16	−5
March 17th–18th	Gallup	*Daily Telegraph*	984	40.5	38.5	18	2
March 17th–18th	NOP	*Independent*	1262	38	42	17	−4
March 18th–20th	MORI	*Sunday Times* (P2)	1257	38	41	19	−3
March 19th–20th	Harris	*Observer*	1096	40	39	17	1
March 20th	ICM	*Sunday Express*	1115	37	42	16	−5
March 19th–21st	NOP	*Independent on Sunday* (P2T)	1004	39	41	15	−2
March 20th–21st	NOP	*Mail on Sunday*	1085	38	40	16	−2
March 21st–23rd	Harris	*Daily Express*	1000	43	38	15	5
March 22nd–23rd	Harris	Independent Television News	2158	38	42	16	−4
March 23rd	MORI	*Times*	1109	38	41	17	−3
March 23rd–24th	Gallup	*Daily Telegraph*	1092	40	40.5	16.5	−0.5
March 24th	ICM	*Guardian*	1096	39	40	17	−1
March 24th	NMR	*European*	1105	38	39	19	−1
March 24th–25th	NOP	*Independent*	1326	39	42	14	−3
March 25th–27th	MORI	*Sunday Times* (P3)	1292	38	40	20	−2
March 26th–27th	Harris	*Observer*	1057	40	38	17	2
March 27th	ICM	*Sunday Express*	1136	36	38	20	−2
March 27th–28th	NOP	*Mail on Sunday*	1099	37	41	18	−4
March 26th–29th	NOP	*Independent on Sunday* (P3T)	1000	39	40	16	−1
March 28th–30th	Harris	*Daily Express*	1108	40	39	17	1
March 29th–30th	Harris	Independent Television News	2152	35	41	19	−6
March 30th	MORI	*Times*	1080	35	42	19	−7
March 31st	ICM	*Guardian*	1126	37	41	18	−4
March 31st–April 1st	Gallup	*Daily Telegraph*	1095	38	37.5	20.5	0.5
March 31st–April 1st	NOP	*Independent*	1302	37	39	19	−2
March 31st–April 3rd	ICM	Press Association	10460	36	39	20	−3
April 1st–3rd	MORI	*Sunday Times* (P4)	1265	37	39	21	−2
April 2nd–3rd	Gallup	*Sunday Telegraph*	1043	37.5	37.5	22	0
April 2nd–3rd	NOP	*Independent on Sunday* (P4T)	1006	38	41	17	−3
April 3rd	ICM	*Sunday Express*	1139	37	39	18	−2
April 3rd–4th	NOP	*Mail on Sunday*	1104	35	41	20	−6
April 3rd–4th	Harris	*Observer*	1090	38	40	17	−2
April 4th–6th	Harris	*Daily Express*	1093	37	38	21	−1
April 4th–7th	Harris	Independent Television News	2210	38	40	18	−2
April 6th–7th	MORI	Yorkshire Television	1065	37	40	20	−3
April 7th–8th	MORI	*Times*	1731	38	39	20	−1
April 7th–8th	NOP	*Independent*	1746	39	42	17	−3
April 7th–8th	Gallup	*Daily Telegraph*	2478	38.5	38	20	0.5
April 8th	ICM	*Guardian*	2186	38	38	20	0

†Source: The Times (1992).
‡NOP, National Opinion Polls; MORI, Market & Opinion Research International; ICM, International Communications and Marketing; P, panel; T, telephone.

members. The MRS reacted quickly to the disaster by convening a group of experts to look at every aspect of the subject. The final report was produced under the chairmanship of David Butler, a well-known and independent-minded academic political scientist, and was published in July 1994 (Market Research Society Working Party, 1994). The MRS report is well researched and well argued and the conclusions were generally accepted as offering a reasonable set of explanations for the failure of the polls. Analysts differ about the relative importance to place on various points, but all accept the broad thrust of the conclusions.

2. CONCLUSIONS OF MARKET RESEARCH SOCIETY REPORT

The first conclusion of the MRS report was that the media and the public expect much greater accuracy from the polls than can be delivered because of the predictive element of opinion polls and the immutable statistical margins of error due to sampling. Nevertheless, the report says, the gap between the poll's findings and the final result in 1992 was greater than their consumers were entitled to expect; indeed, it was greater than had ever been delivered since polls began. The report then explored the factors that may have contributed to this gap and concluded that the main factors were

(a) a late swing to the Conservatives,
(b) differential response by party, implying that Conservatives were more likely to give wrong responses,
(c) differential non-response by party, implying that Conservatives were more likely to refuse, and
(d) an effect due to quota sampling.

All these factors, if corrected, would tend to increase the proportion of Conservatives. In addition several minor factors were identified, and again on this occasion they all tended to bias the results against the Conservatives. In the context of the 1992 election all the potential biases appear to have operated in the same direction at the same time thus causing the disaster. The implication is that in previous elections biases may fortuitously have cancelled one another out, leading to a satisfactory outcome. Biases are not random variables; they are context specific, and the art is to identify the causes and the context that determines their values. The failure in 1992 does not imply that at the next election the polls will fail, any more than the successes in earlier elections implied success in 1992. However, one aspect of the polls of particular interest to statisticians is that of quota sampling, and this is now examined in greater detail.

3. QUOTA SAMPLING

The MRS report identified aspects of quota sampling as contributory factors to the bias. To quote the report:

> 'Some inadequacies were revealed in the operation of the quota system to select representative samples. This arose partly because quotas and weights did not reflect sufficiently accurately the social profile of the electorate, and partly because the variables used as the basis of quotas and in corrective weighting were not closely enough correlated

with voting behaviour to ensure that the samples were fully representative of the distribution of political support amongst the electorate.'

Note that it is not the quota method that is blamed but its implementation in this particular year. The evidence for selection bias varies with company (see Table 2), and yet there is no evidence of a difference between companies which cannot be explained by the random choice of sampling points at the first stage of sampling. This is one aspect of the MRS report that deserves further examination, but before attempting this it is worth looking at the historical record of the polls.

Since 1979 none of the polls have used any form of random sampling for the selection of the final units, the potential voters. However, all the polls stratify the parliamentary constituencies, which constitute the primary sampling units (PSUs) or sampling points, and select a master sample of constituencies by using rules similar to those in two-stage random sampling. Deviations from random sampling occur at the second stage when the interviewers have freedom to choose respondents within the constraints of their quotas. Fig. 1 shows the historical record of the polls since 1964, when both random and quota samples were carried out simultaneously. It is clear from Fig. 1 that the record of random polls is no better, and indeed may even be

TABLE 2
Proportion of council tenants measured in 22 campaign polls†

Pollster‡	*Poll*	*Unweighted (%)*	*Weighted (%)*
Census, 1991			19.5
National Readership Survey, 1991			21.5
NOP	*Mail on Sunday* 2	17.5	24.0§
NOP	*Independent* 3	18.8	24.0§
Gallup	Final poll	19.0	20.0
NOP	*Mail on Sunday* 4	19.7	24.0§
Gallup	Second poll	20.0	21.0
Gallup	Third poll	20.0	20.0
NOP	*Independent* 1	20.0	24.0§
NOP	*Mail on Sunday* 3	20.4	24.0§
NOP	*Independent* 2	20.5	24.0§
MORI	*Times* final poll	20.6	24.0§
NOP	*Independent* final poll	20.7	24.0§
MORI	*Times* fourth poll	21.6	22.9§
ICM	Press Association 10,000 poll	22.0	23.0
ICM	Final poll	22.0	22.0
MORI	*Times* second poll	22.4	22.5§
Gallup	First poll	23.0	24.0
Gallup	Fourth poll	23.0	23.0
Harris	Independent Television News second poll	23.0	17.0§
MORI	*Times* third poll	23.2	23.4§
MORI	*Times* first poll	23.2	23.2
Harris	Independent Television News final poll	25.0	25.0
Harris	Independent Television News first poll	26.0	25.0

†Source: Market Research Society Working Party (1994).
‡NOP, National Opinion Polls; MORI, Market & Opinion Research International; ICM, International Communications and Marketing.
§Polls weighted by tenure.

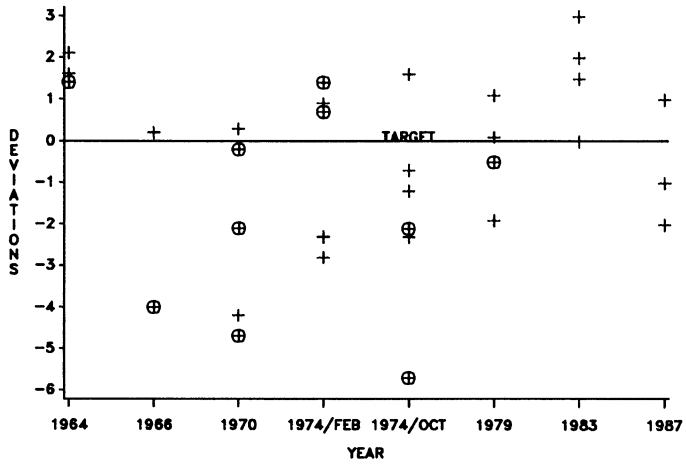

Fig. 1. Historical record of the final polls and election results, Conservative votes: +, quota samples; ⊕, random samples

worse, than that of quota polls during the period from 1964 to 1979. On the basis of this empirical evidence, and given the pressure from the media for timely yet inexpensive results, the polling companies have abandoned random sampling for opinion polls since 1979. There were no serious problems in 1983 and 1987, nothing to indicate the disaster of 1992. Thus explanations of the disaster depend almost exclusively on the benefits of hindsight.

When polls fail a frequent excuse is that there has been a late swing in support from one party to another, so late that the polls could not detect it. In the UK this has led to polls being conducted as late as possible, sometimes on the day before the election. One way of examining the trend during the campaign is to ignore the polls in the month before the election and to make a prediction of the election result on the basis of the time series of regular monthly polls. Using the methods in Scott *et al.* (1977) and Smith (1978), an ARMA(1, 1) model was fitted to the monthly Gallup poll data showing the Conservative lead over Labour for the periods from January 1979 to June 1983 (Fig. 2), to June 1987 (Fig. 3) and to April 1992 (Fig. 4), before the elections in those years. The one-month-ahead forecasts show that in 1983 and 1987 the forecasts were well within the range of prediction error, which itself must be greater than the implicit sampling error. This suggests that in both elections the campaign had little effect. The plot of the data before the 1992 election is fascinating; the results for almost a year before the election are all consistent with the final polls before the election and forecast the final poll outcomes accurately. Unfortunately the polls failed to predict the actual result. Whatever the causes of this failure they appear to have been operating consistently for many months before the 1992 election. Thus any explanation based on short-term effects is not consistent with the data; it is the long-term context that matters.

The feature of quota sampling that is of most concern to statisticians is the treatment of non-response. None of the polls declared the level of non-response, which for interviews in the street is almost impossible to determine. Quota samplers argue that the quota controls act to minimize the effect of non-response, but there is

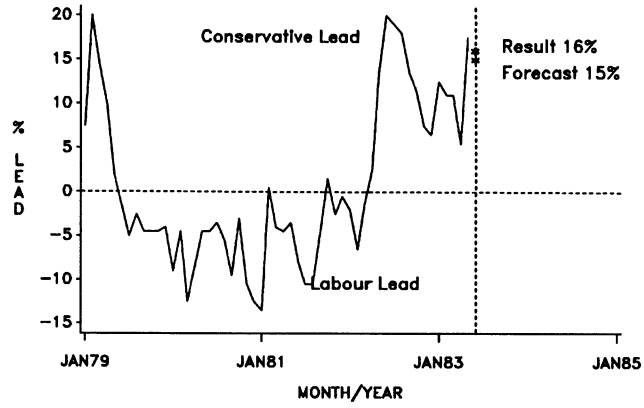

Fig. 2. Conservative lead, January 1979–June 1983: ARMA(1, 1) forecast (source: Gallup)

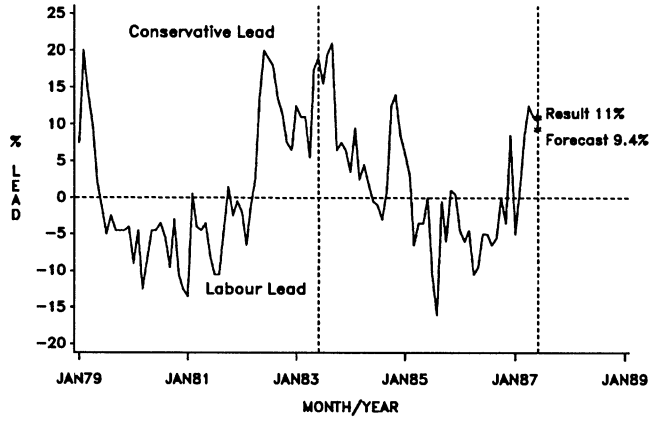

Fig. 3. Conservative lead, January 1979–June 1987: ARMA(1, 1) forecast (source: Gallup)

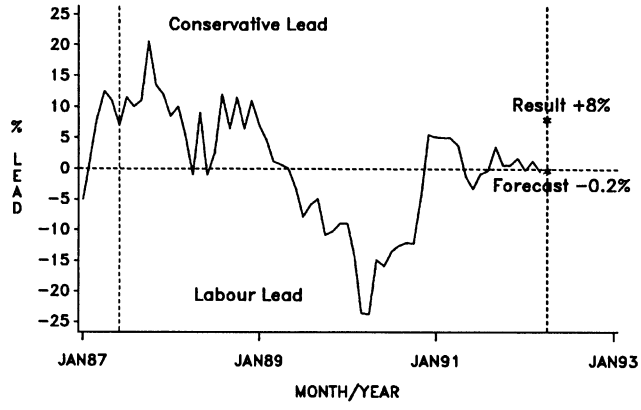

Fig. 4. Conservative lead, January 1979–April 1992: ARMA(1, 1) forecast (source: Gallup)

little evidence to support this assertion. Random samplers are very critical of this aspect of quota sampling and Jowell *et al.* (1994) have suggested that for quota samples based on interviews at addresses the respondents should be similar to those who respond at the first wave of interviews in a random survey with call-backs. Fortunately an academic random survey, the British Election Survey, was conducted at the same time as the polls, and this could be used to see whether there was an availability bias in the first wave of responses. The data showed a strong bias in favour of the Labour Party in the first wave and this could explain as much as half the bias observed in 1992. The MRS report considered this study and concluded that the effect was exaggerated by the weighting used, was not necessarily relevant to street interviews and failed to take into account the time-of-day effect which is controlled in some quota samples. Even on their analysis availability bias could account for a quarter of the total bias. The Jowell *et al.* (1994) study makes a useful contribution to the debate and shows the benefit of conducting a well-designed random survey alongside the commercial polls for evaluation and validation.

Quota sampling is a form of balanced sampling where the configuration of the sample matches that in the population for certain key variables. In 1992 the results of the 1991 census were not yet available and so the polls used weights based on a large random sample, the National Readership Survey. It transpired that these weights were inaccurate relative to those of the census, and again the biases all seemed to favour Labour. The MRS report concluded that far more attention should be paid to the weights and to the choice of quota variables. If weighting is to be used either implicitly or explicitly then the weights must be valid weights.

The main conclusion is that in 1992 a whole series of sources of bias conspired to work in the same direction thus creating a large overall bias. It is not possible to generalize from this to future election polls since the conditions that led to the failure in 1992 may not be repeated. Many of the explanations of the causes of bias have been based on studies carried out alongside the polls, not on data from the polls themselves. In fact it is a truism that biases can only be determined from external evidence, and the lesson for future polls is that a well-designed and well-funded random panel survey should be carried out alongside the polls to determine the effects operating in the context of the particular election. A more detailed proposal along these lines is made by Lynn and Jowell (1996).

4. VARIABILITY OF QUOTA SAMPLES

As stated in Section 1 a simple sign test is sufficient to demonstrate that the problem with the polls in 1992 was bias not variance. However, the MRS report does make the point that sampling variability is an immutable source of statistical error, and one of the problems with quota samples is that there is no standard procedure for estimating sampling variances. The replication implicit in the 50 or so polls carried out in one month provides a unique opportunity to estimate the quota sampling variances empirically. Collins (1988) carried out an analysis of the variation between the polls before the 1987 election and concluded that the quota sampling variance was consistent with that of a two-stage random sample of the same size. His analysis was based on the differences between polls from the same company over time and between different companies at the same time and over different times. He could ignore the time dimension because there was little evidence of any effect on

anticipated voting in the month before the election. In a similar study based on the 1992 polls Lynn (1994) showed that similar results held in 1992. It thus appears that the variances of quota samples are similar to those of two-stage random samples, namely slightly bigger than those for a simple random sample of the same size, for the variables measured in political opinion polls.

All the polling companies use a two-stage design based on a master sample of constituencies, but there is variation in the number of PSUs chosen by each company and in the number of interviews in each PSU. The estimated variances should take into account these different allocations. For comparisons within companies the sample of first-stage units is held constant, whereas for comparisons between companies there is variation between the selected PSUs, and in the number of PSUs selected. None of these features has been taken into account in the previous variance studies.

For a subset of 32 of the 50 polls listed in Table 1, it was possible to obtain not only the estimate p_{ti} for company i in week t but also the total sample size n_{ti} for each poll and the number of sampling points (PSUs), m_{ti}, used on each occasion. The data

TABLE 3
Company-specific variances for the proportion voting Conservative

Case	*Company*	*Week*	*n*	*m*	*Specific variance*	$p_C = 39\%$ *simple random sample variance*
1	1	1	1050	54	3.20	2.27
2	1	2	1085	54	3.15	2.19
3	1	3	1099	54	3.14	2.16
4	1	4	1104	54	3.13	2.15
5	2	1	1054	53	3.23	2.26
6	2	1	1099	54	3.14	2.16
7	2	2	1109	55	3.09	2.15
8	2	3	1080	54	3.16	2.20
9	3	4	1731	164	1.45	1.37
10	4	1	1054	110	2.29	2.26
11	4	2	1096	110	2.23	2.17
12	4	3	1057	110	2.28	2.25
13	4	4	1090	110	2.24	2.18
14	5	1	1086	110	2.25	2.19
15	5	1	1081	110	2.25	2.20
16	5	2	1000	110	2.36	2.38
17	5	3	1108	110	2.22	2.15
18	5	4	1093	110	2.24	2.18
19	6	1	984	100	2.48	2.42
20	6	2	1092	110	2.24	2.18
21	6	3	1095	110	2.23	2.17
22	6	4	1043	105	2.34	2.28
23	7	4	2478	250	0.99	0.96
24	8	1	1100	52	3.20	2.16
25	8	2	1096	52	3.21	2.17
26	8	3	1126	52	3.17	2.11
27	9	4	2186	103	1.62	1.09
28	10	1	1059	52	3.26	2.25
29	10	2	1115	52	3.19	2.13
30	10	3	1136	52	3.16	2.09
31	10	4	1139	52	3.16	2.09
32	11	4	10460	330	0.43	0.23

used for the variance study are listed in Table 3. This information allows the sampling variance to be modelled and the variance to be partitioned between both stages of sampling so that company-specific variances can be estimated. We assume that the quota sampling variance can be expressed in the same form as the variance of a two-stage sample, with contributions from both stages of sampling. Specifically we ignore finite population terms and assume that the variance of an estimator of a proportion, or of linear functions of proportions, can be written as

$$V(p_{ti}) = \frac{\sigma_1^2}{m_{ti}} + \frac{\sigma_2^2}{n_{ti}}, \tag{4.1}$$

where p_{ti} is the quota sample estimator of the population proportion p_t, σ_1^2 is the variance due to the first stage of sampling and σ_2^2 is the variance due to the second stage of sampling. With this structure for the variance function we can write down a model for each estimate of the form

$$p_{ti} = p_t + b_{1ti}e_{1i} + b_{2ti}e_{2ti}, \tag{4.2}$$

where t indexes the week of the poll, i indexes the company carrying out the poll and

$$E(e_{1i}) = E(e_{2ti}) = 0,$$
$$V(e_{1i}) = \sigma_1^2, \qquad V(e_{2ti}) = \sigma_2^2,$$
$$b_{1ti} = m_{ti}^{-1/2}, \qquad b_{2ti} = n_{ti}^{-1/2},$$

and all the covariances between the errors are assumed to be 0.

Model (4.2) was fitted to the data in Table 3 by using the iterative generalized least squares algorithm employed in the multilevel model program ML3; see Goldstein (1995). Note that the first stage of sampling corresponds to level 2 in a multilevel model. Variances were estimated for the proportions voting Conservative, p_C, Labour, p_L, Liberal Democrat, p_{LD}, and for the Conservative lead, $p_C - p_L$. Estimates of the variances are shown in Table 4. Although the first-stage variances do not appear to be significantly different from 0, the structure of the sample design expressed through model (4.2) requires that both terms be included. This is not an exercise in model choice but in variance estimation, and statistical significance is simply a surrogate for sample size.

The primary objective in this analysis is not to model the structure of the sample estimates of the proportions p_{ti} but to estimate the structure of the survey variance function. The survey errors are measured on the scale of the original proportions and

TABLE 4

Estimates of components of variance: different means each week†

Proportion estimated	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$
p_L	45.0 (37.8)	1691 (508.7)
p_C	102.2 (60.0)	1317 (402.2)
p_{LD}	26.9 (26.8)	1439 (430.4)
$p_C - p_L$	173.4 (131.0)	5039 (1522.0)

†Standard errors are given in parentheses.

so the usual logistic transformations used in a generalized linear model analysis are not appropriate in this study. Since the proportions are approximately constant during the period under study, any modelling of the proportions, which would be of substantive rather than statistical interest, would not require a transformation. The estimates of the proportions for each week are shown in Table 5. There is some evidence from the values in Table 5 that there was a change in anticipated voting after the first two weeks. The proportions expecting to vote for the Liberal Democrats increased at the expense of the other two parties, whereas the lead is unaffected. This conjecture can be tested by fitting model (4.2) with the first two proportions and the last two proportions constrained to be equal. Assuming normality for the distribution of the proportions a likelihood ratio test confirms that the constrained model is a good fit. Further constraining the model to a single value, suggesting no effect of the campaign during the election period, gives a poor fit except for the Conservative lead. These results are consistent with those of Clifford and Heath (1994) who fitted models to the complete set of poll results, but without taking into account the variance structure. Clifford and Heath (1994) found a significant company effect, which in our model will be reflected in the first-stage component of variance, due to the use of master samples. Explicit modelling of the sampling variance would seem to have advantages in explaining these results.

The estimates of the components of sampling variance are affected by the assumptions made about the proportions. In Table 6 the components of variance under the model with fixed means for the first two and the last two weeks are shown. Comparing these results with those in Table 4 shows that the changes are relatively small. One advantage of estimating the components of variance is that company-specific sampling variances can be estimated which take into account the number of

TABLE 5
Estimates of the proportions voting for each party by week

Week	*Proportion voting (%)*			
	Conservative, p_C	*Labour,* p_L	*Liberal Democrat,* p_{LD}	*Conservative lead* $p_C - p_L$
1	39.8	40.5	15.8	−1.5
2	39.6	39.8	16.4	−0.3
3	38.0	39.2	18.5	−1.4
4	37.0	38.8	19.9	−1.9

TABLE 6
Estimates of components of variance: first two weeks' and last two weeks' proportions equal†

Proportions estimated	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$
p_L	49.2 (40.4)	1751 (527.2)
p_C	96.2 (59.2)	1489 (435.5)
p_{LD}	29.5 (31.5)	1804 (538.6)
$p_C - p_L$	159.4 (127.9)	5384 (1622.0)

†Standard errors are given in parentheses.

PSUs selected and the interviewer workloads within a PSU. Previous studies of quota sampling variances have assumed homogeneity across companies. Using the results from Table 6 the company-specific variances for the proportion voting Conservative are shown in Table 3, together with the number of PSUs and the total sample size. Also shown are the simple random sampling variances based on $p_C = 39\%$.

Comparing case 1 with case 10 it can be seen that for the same total sample size it is better to select more PSUs and to have fewer interviews per PSU. This accords with the results from random sampling. Using the results from Table 3 it is possible to calculate design effects for each case. Assuming that $p_C = 39.0\%$ the simple random sampling variance for case 1 is 2.27, giving a design effect of 1.41. For case 10 the corresponding design effect is 1.01. For case 9 the design effect is 1.06, and for case 32 it is 1.87. These results suggest that it is not safe to assume that the polls of the same size are homogeneous with respect to sampling variation.

Results have also been obtained for the other proportions and for the Conservative lead. For case 1 the specific variance for the lead is 8.08, which demonstrates the effect of the negative correlation between multinomial proportions on variances. The simple random sampling variance is 7.53, giving a design effect of 1.07. The design effect on the difference is much smaller than that on the individual proportions, and again this is in line with expectation.

Overall the results obtained by modelling the variances have been consistent with previous analyses, which suggests that the methodology is sound, but this model has the great advantage of allowing company-specific variances to be estimated. There are great potential statistical benefits in analysing natural experiments with a strong element of replication whenever the opportunity arises, and for concentrating on the analysis and modelling of the structure of the immutable random variation.

ACKNOWLEDGEMENTS

This research was supported by a grant from the Economic and Social Research Council of the UK under the Analysis of Large and Complex Datasets Programme. Denise Silva and Fiona Steele provided valuable computing assistance.

REFERENCES

Clifford, P. and Heath, A. (1994) The election campaign. In *Labour's Last Chance?* (eds A. F. Heath, R. Jowell, J. Curtice and B. Taylor). Aldershot: Dartmouth Press.

Collins, M. (1988) Lessons from the polls. *Market Research Society Conf., Brighton, Mar. 16th–18th.*

Goldstein, H. (1995) *Multilevel Statistical Models*, 2nd edn. London: Arnold.

Jowell, R., Hedges, B., Lynn, P., Farrant, G. and Heath, A. (1993) The 1992 British Election: the failure of the polls. *Publ. Opin. Q.*, **57**, 238–263.

Lynn, P. (1994) Quota sampling variance.

Lynn, P. and Jowell, R. (1996) How might opinion polls be improved?: the case for probability sampling. *J. R. Statist. Soc.* A, **159**, 21–28.

Market Research Society Working Party (1994) *The Opinion Polls and the 1992 General Election.* London: Market Research Society.

Scott, A. J., Smith, T. M. F. and Jones, R. G. (1977) The application of time series methods to the analysis of repeated surveys. *Int. Statist. Rev.*, **45**, 13–28.

Smith, T. M. F. (1978) Principles and problems in the analysis of repeated surveys. In *Survey Sampling and Measurement* (ed. N. K. Namboodiri), pp. 201–216. New York: Academic Press.

The Times (1992) *Times Guide to the House of Commons.* London: The Times.