CONTENTS

1. INTRODUCTION

Consider a pair of successive elections of the British type, where the population is divided into constituencies and each constituency returns one elected member. In any one constituency we know, for each election, the number of people who voted for each party and the number that did not vote. What we do not know, however, is how individual electors changed their vote between the two elections, and this would clearly be of interest to students of political behaviour. For example, if the Liberal vote is seen to have increased at the second election, one might ask how much of the Liberal Party's new support came from each of the other parties and how much from abstainers at the first election. Under secret ballot arrangements, this type of question cannot be answered precisely. One might attempt to provide answers by conducting some kind of sample survey where people are asked to recall their votes at the two elections, or even using a "panel" survey where a group of electors is followed through the two elections and at each election is asked how they voted. However, it has been shown that the results of this type of survey are unreliable and prone to bias; see, for example, the discussion by Miller (1972).

The only alternative, then, is to use the voting totals in each constituency, which are aggregates of the quantities in which we are interested, to estimate the pattern of individual voting change. To do this we need the election results from several constituencies and we must make some assumptions about the inter-constituency structure of individual voting change; for example, we might assume that the pattern of change is the same in all constituencies, or that it depends on the strength of the Liberal Party in a constituency, or on some other constituency-related variable. Estimation of models of this type has been investigated by Hawkes (1969), Irwin and Meeter (1969), Miller (1972) and McCarthy and Ryan (1977). The aim of this project is to further investigate methods of timation, and to examine the possibility of a "random effects" model account for variation in the pattern of change between constituencies.

2. VOTER TRANSITIONS : THE MODELS TO BE CONSIDERED

GENERAL FORMULATION

Suppose that there are

K constituencies, each of which has the same

I parties at the first election (including an imaginary "abstainers" party) and

J parties at the second election (again including an "abstainers" party)

(NOTE: It is not necessary that I and J are equal; the parties themselves may even be different at the two elections, although it is necessary that, at any one of the two elections, the party line-up is the same in all K constituencies.)

Let Z_{ik}, $1 \leq i \leq I$, $1 \leq k \leq K$, be the number of people in constituency k voting for party i at the first election; and let Y_{jk}, $1 \leq j \leq J$, $1 \leq k \leq K$, be the number of people in constituency k voting for party j at the second election.

Of those in constituency k who voted for party i at the first election, suppose that a proportion a_{ijk} subsequently vote for candidate j at the second election; then we can write $Y_{jk} = \sum_{i=1}^{I} a_{ijk} Z_{ik}$, $j=1,\ldots,J$.

This, of course, assumes that the electorate is stable - an assumption which will not be true in general, due to death, migration and coming of age. This may be overcome by supposing that:

"At the time of the second election, the proportion of the electorate who would have voted for candidate i at the last election, if they had been able to, is the same as the proportion who actually did so at the time." (2.1)*

(* An assumption first stated by Hawkes (1969))

This seems a reasonable supposition to make as long as the two elections not too far separated in time and the franchise rules remain unchanged. consequence of assuming (2.1) is that we merely scale the original Z_{ik} make the total electorate in constituency k have the same size, i.e. replace Z_{ik} by

$$X_{ik} = Z_{ik}\left(\sum_{j=1}^{J} Y_{jk} \Big/ \sum_{i=1}^{I} Z_{ik}\right) \qquad (2.2)$$

hen

$$Y_{jk} = \sum_{i=1}^{I} a_{ijk}X_{jk} = \sum_{i=1}^{I} T_{ijk}\,, \quad \text{say,} \qquad (2.3)$$

ubject to the condition

$$\sum_{j=1}^{J} a_{ijk} = 1 \qquad \forall\ i,k \qquad (2.4)$$

In (2.3), T_{ijk} is the actual number of transitions from i to j in constituency k. This may be represented in a transition table:

	Y_{1k}	Y_{Jk}
X_{1k}	T_{11k}	T_{1Jk}
.	.	.
.	.	.
.	.	.
.	.	.
.	.	.
.	.	.
X_{Ik}	T_{I1k}	T_{IJk}

(2.5)

This is the table for constituency k; there are K such tables in all.

2.2 MODEL A : Fixed multinomial rows

This model supposes that for each constituency, k, given the results $(X_{1k},\ldots,X_{Ik})$ of the first election, row i of the transition table (2.5) has a multinomial distribution with

$$\begin{cases} \text{index } X_{jk} \\ \text{parameters } (\Pi_{i1},\Pi_{i2},\ldots,\Pi_{iJ}) \end{cases}$$

I rows being distributed independently.

we have an I x J matrix

$$\Pi = \begin{pmatrix} \Pi_{11} & \Pi_{12} & \cdots & \Pi_{1J} \\ \vdots & & & \vdots \\ \Pi_{I1} & \cdots & \cdots & \Pi_{IJ} \end{pmatrix}$$

transition matrix, assumed constant over the K constituencies, Π_{ij} ($1 \leq i \leq I$, $1 \leq j \leq J$) is the probability that an elector who voted party i at the first election will vote for party j at the second.

Since the quantities Π_{ij} are probabilities, the following must hold:

$$0 \leq \Pi_{ij} \leq 1 \qquad (1 \leq i \leq I,\ 1 \leq j \leq J) \tag{2.6}$$

and

$$\sum_{j=1}^{J} \Pi_{ij} = 1 \qquad (1 \leq i \leq I) \tag{2.7}$$

The equations (2.7) imply that there are in fact only $I \times (J-1)$ free meters, and the estimation problem for this model is concerned with e $I \times (J-1)$ parameters.

MODEL B : "Random" multinomial rows

Here we again suppose that row i of the transition table (2.5) for stituency k has a multinomial distribution, this time with

$$\begin{cases} \text{index } X_{ik} \\ \text{parameters } (\Pi_{i1k}, \Pi_{i2k}, \ldots, \Pi_{iJk}) \end{cases}$$

i.e. the transition probabilities are allowed to be different in fferent constituencies. This model then assumes that the J-tuples

$$(\Pi_{i1k}, \Pi_{i2k}, \ldots, \Pi_{iJk}) \qquad (1 \leq i \leq I)$$

each of the K constituencies are drawn independently from a Dirichlet stribution with parameters $(\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{iJ})$ $(1 \leq i \leq I)$

$$\text{with} \quad \alpha_{ij} \geq 0 \quad \forall\ i,j \tag{2.8}$$

That is, in each constituency k, the transition vector $(\Pi_{i1k}, \Pi_{i2k}, \ldots, \Pi_{iJk})$

for party i has probability density function

$$f(\Pi_{i1k},\Pi_{i2k},\dots,\Pi_{iJk}) = \frac{\Gamma(\alpha_{i1}+\dots+\alpha_{iJ})}{\Gamma(\alpha_{i1})\Gamma(\alpha_{i2})\dots\Gamma(\alpha_{iJ})}\Pi_{i1k}^{\alpha_{i1}-1}\Pi_{i2k}^{\alpha_{i2}-1}\dots\Pi_{iJk}^{\alpha_{iJ}-1} \qquad (2.9)$$

Thus, under this model, each row of the transition table (2.5) for a particular constituency has a Dirichlet-compound-multinomial distribution. The parameters of interest are now $\{\alpha_{ij} : 1 \leq i \leq I,\ 1 \leq j \leq J\}$, the parameters of the I underlying Dirichlet distributions, which are assumed constant over the K constituencies; so there are now $I \times J$ parameters to be estimated.

2.4 THE ADVANTAGE OF MODEL B

Model A, with the same transition probabilities assumed for each constituency, has been found to fit data from British General Elections very badly (see, for example, Hawkes (1969)); it has not been possible to find a reasonably large group of K constituencies with sufficient homogeneity to allow the same transition matrix to describe the changes occurring in all of the constituencies.

Model B, however, allows some variation in the transition matrix between constituencies, so it might be expected to fit the data rather better. Moreover, the re-parameterisation from

$$\Pi = \begin{pmatrix} \Pi_{11} & \dots & \Pi_{IJ} \\ \vdots & & \vdots \\ \Pi_{I1} & \dots & \Pi_{IJ} \end{pmatrix} \quad \text{with I constraints on the parameters} \quad \text{(Model A)}$$

to

$$\alpha = \begin{pmatrix} \alpha_{11} & \dots & \alpha_{1J} \\ \vdots & & \vdots \\ \alpha_{I1} & \dots & \alpha_{IJ} \end{pmatrix} \quad \text{with all parameters free} \quad \text{(Model B)}$$

does not cause any difficulty in interpretation; for under Model B,

$$E(\Pi_{ijk}) = \frac{\alpha_{ij}}{\sum_{m=1}^{J}\alpha_{im}} = p_{ij}, \text{ say } \begin{pmatrix} 1 \leq i \leq I \\ 1 \leq j \leq J \end{pmatrix} \qquad (2.10)*$$

(*See, for example, Wilks (1962) for properties of the Dirichlet distribution)

$$\text{and Var } (\Pi_{ijk}) = \frac{\alpha_{ij}(\sum_{m=1}^{J}\alpha_{im} - \alpha_{ij})}{(\sum_{m=1}^{J}\alpha_{im})^2 (\sum_{m=1}^{J}\alpha_{im} + 1)} \qquad (2.11)*$$

So the ratios $\{p_{ij} : 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J\}$ are the "average" transition probabilities across the K constituencies, whilst the sum $A_i = \sum_{m=1}^{J} \alpha_{im}$ gives a good indication of the amount of variation in transition probabilities between constituencies; the smaller the value of A_i, the more variation there is.

Despite the shortcomings of Model A, it will still be useful to estimate its parameters for purposes of comparison.

2.5 BRIEF OUTLINE OF THE TWO ESTIMATION PROCEDURES TO BE INVESTIGATED

1. Maximum Likelihood, based on a multivariate normal approximation to the distribution of the vectors $\underset{\sim}{Y}_k$ of results of the second election.

This can be used to estimate the parameters of both models A and B.

Why use an approximation? Take, for example, model A; in each constituency the rows of transitio n table (2.5) are multinomial, with probability distribution

$$\Pr\ \{T_{i1k}, T_{i2k}, \ldots, T_{iJk} | X_{ik}\} = X_{ik}! \prod_{j=1}^{J} \left(\frac{\Pi_{ij}^{T_{ijk}}}{T_{ijk}!} \right) \qquad (2.12)$$

The I rows of the transition table are independently distributed, so the probability of any particular transition table is

$$\Pr\{T_{ijk} : 1 \leqslant j \leqslant I,\ 1 \leqslant j \leqslant J | X_k\} = \prod_{i=1}^{I} \left(X_{ik}! \prod_{j=1}^{J} \left(\frac{\Pi_{ij}^{T_{ijk}}}{T_{ijk}!} \right) \right) \qquad (2.13)$$

Let $\mathcal{T}_k$ be the set of all possible transition tables that could have resulted in the observed election totals $\underset{\sim}{X}_k$ and $\underset{\sim}{Y}_k$. Then

$$\Pr(\underset{\sim}{Y}_k | \underset{\sim}{X}_k) = \sum_{\mathcal{T}_k} \prod_{i=1}^{I} \left(X_{ik}! \prod_{j=1}^{J} \left(\frac{\Pi_{ij}^{T_{ijk}}}{T_{ijk}!} \right) \right) \qquad (2.14)$$

(*See, for example, Wilks (1962) for properties of the Dirichlet distribution)

Now constituencies behave independently, so the likelihood function r all K constituencies is

$$L(\Pi) = \prod_{k=1}^{K} \left(\sum_{T_k} \prod_{i=1}^{I} \left(X_{ik}! \prod_{j=1}^{J} \left(\frac{\Pi_{ij}^{T_{ijk}}}{T_{ijk}!} \right) \right) \right) \qquad (2.15)$$

a product of a sum of products. It would, in theory, be possible to rive maximum likelihood estimates from this exact likelihood, but the ount of computation involved would be enormous.

So we seek a suitable approximation to the likelihood; the vote unts Y_{jk} may be thought of as the sum of a very large number of variables, ch of which takes the value 0 or 1, and so by central limit theory are kely to have a distribution that is approximately normal. Therefore we proximate the distribution of the vectors $\underset{\sim}{Y}_k$ with a multivariate normal stribution, and maximise the likelihood given by that approximation.

Details of the method are given in Chapter 3.

2. <u>The Expectation-Maximisation (E-M) Algorithm</u> : Dempster, Laird and bin (1977) describe the E-M algorithm as "a broadly applicable algorithm r computing maximum likelihood estimates from incomplete data". The timation of transition matrices from election totals can be regarded as "incomplete data" problem; for the "complete" data here would be the ternal elements T_{ijk} of the transition table for each constituency, from ich it would be a simple matter, at least in the cast of model A, to obtain ximum likelihood estimates for the elements of the matrix Π. We have at r disposal only aggregate, or "incomplete" data, namely the sums $Y_{jk} = \sum_{i=1}^{I} T_{ijk}$.

In this particular situation, estimating the parameters Π of model A, e E-M algorithm proceeds as follows:

(i) Start with an initial estimate of Π

<u>(E-step)</u> (ii) Calculate $T_{ijk} = E\,(T_{ijk} \mid X_k, Y_k, \Pi)$ $\quad i,j,k$

(iii) Re-estimate Π by maximum likelihood, treating the

<u>(M-step)</u> T_{ijk} as if they were the true entries in the transition table.

(iv) Check for convergence of the sequence of Π's and return to (ii) if not converged.

Details of the method are given in Chapter 4.

3. MAXIMUM LIKELIHOOD ESTIMATION BASED ON A NORMAL APPROXIMATION

3.1 THE NORMAL APPROXIMATION

We have $Y_{jk} = \sum_{i=1}^{I} T_{ijk}$ $\quad (1 \leq j \leq J,\ 1 \leq k \leq K)$ (3.1)

Under model A:

$$E(Y_{jk}) = \sum_i X_{ik}\Pi_{ij} \tag{3.2}$$

$$\mathrm{var}(Y_{jk}) = \sum_i \mathrm{var}(T_{ijk})$$

$$= \sum_i X_{ik}\Pi_{ij}(1-\Pi_{ij}) \tag{3.3}$$

and $\mathrm{cov}\,(Y_{jk}, Y_{j'k}) = \mathrm{cov}(\sum_i T_{ijk}, \sum_i T_{ij'k})$

$$= \sum_i \mathrm{cov}\,(T_{ijk}, T_{ij'k})$$

$$= -\sum_i X_{ik}\Pi_{ij}\Pi_{ij'} \qquad (j \neq j') \tag{3.4}$$

Under model B:

As in (2.10) write $p_{ij} = \alpha_{ij}/\sum_j \alpha_{ij}$ $\quad (1 \leq i \leq I,\ 1 \leq j \leq J)$

Then

$$E(Y_{jk}) = \sum_i X_{ik}\, p_{ij} \tag{3.5}$$

$$\mathrm{var}(Y_{jk}) = \sum_i X_{ik}\left(\frac{X_{ik} + \sum_j \alpha_{ij}}{1 + \sum_j \alpha_{ij}}\right) p_{ij}(1-p_{ij}) \tag{3.6}$$

and $$\mathrm{cov}\,(Y_{jk}, Y_{j'k}) = -\sum_i X_{ik}\left(\frac{X_{ik} + \sum_j \alpha_{ij}}{1 + \sum_j \alpha_{ij}}\right) p_{ij}p_{ij'} \quad (j \neq j') \tag{3.7}$$

(See, for example, Johnson and Kotz (1969) for properties of the compound multinomial distribution).

If we write

$$P_{ij} = \begin{cases} \Pi_{ij} & \text{(model A)} \\ \alpha_{ij}/\sum_j \alpha_{ij} & \text{(model B)} \end{cases} \qquad \begin{pmatrix} 1 \leq i \leq I \\ 1 \leq j \leq J \end{pmatrix} \tag{3.8}$$

and $$W_{ik} = \begin{cases} 1 & \text{(model A)} \\ \dfrac{X_{ik} + \sum_j \alpha_{ij}}{1 + \sum_j \alpha_{ij}} & \text{(model B)} \end{cases} \qquad \begin{array}{l} 1 \leq i \leq I \\ 1 \leq j \leq J \end{array} \qquad (3.9)$$

then the expectations, variances and covariances have the same form for both models, i.e.

$$E(Y_{jk}) = \sum_i X_{ik} p_{ij} \qquad (3.10)$$

$$\text{var}(Y_{jk}) = \sum_i X_{ik} W_{ik} p_{ij}(1-p_{ij}) \qquad (3.11)$$

$$\text{cov}(Y_{jk}, Y_{j'k}) = -\sum_i X_{ik} W_{ik} p_{ij} p_{ij'} \qquad (j \neq j') \qquad (3.12)$$

Since $\sum_{j=1}^{J} Y_{jk}$ is fixed by the size of the electorate in constituency k, we need only consider the first (J-1) variables; define

$$\underset{\sim}{Y}_k^{*T} = (Y_{1k}, Y_{2k}, \ldots, Y_{(J-1),k}) \qquad (3.13)$$

$$P^* = \begin{pmatrix} p_{11} & \cdots & p_{1,J-1} \\ \vdots & & \vdots \\ p_{I1} & \cdots & p_{I,J-1} \end{pmatrix} \qquad (3.14)$$

and let Γ_k^* be the variance-covariance matrix of $\underset{\sim}{Y}_k^*$. Then from (3.11) and (3.12) we have

$$\Gamma_k^* = \text{Diag}\left[P^{*T}\underset{\sim}{U}_k\right] - P^{*T}\text{Diag}\left[\underset{\sim}{U}_k\right] P^* \qquad (3.15)$$

where $$U_{ik} = X_{ik} W_{ik} \qquad (1 \leq i \leq I,\ 1 \leq k \leq K) \qquad (3.16)$$

Now, from central limit theory, $\underset{\sim}{Y}_k^*$ has (asymptotically) a multivariate normal distribution,

$$\underset{\sim}{Y}_k^* \underset{as}{\longrightarrow} N_{J-1}(P^{*T}\underset{\sim}{X}_k, \Gamma_k^*) \qquad (3.17)$$

The log-likelihood function for this distribution is

$$L(P^*, \underset{\sim}{W}_1, \underset{\sim}{W}_2, \ldots, \underset{\sim}{W}_k \mid \underset{\sim}{Y}_1^*, \underset{\sim}{Y}_2^*, \ldots, \underset{\sim}{Y}_k^*)$$

$$= \frac{K(J-1)}{2} \log(2\Pi) - \frac{1}{2} \sum_k \log\left|\Gamma_k^*\right| - \frac{1}{2} \sum_k (\underset{\sim}{Y}_k^* - P^{*T}\underset{\sim}{X}_k)^T \Gamma_k^{*-1} (\underset{\sim}{Y}_k^* - P^{*T}\underset{\sim}{X}_k) \qquad (3.18)$$

We now wish to maximise this function

(i) over $\{P_{ij}\}$ $(1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J-1)$ in the case of of model A and (ii)

over $\{\alpha_{ij}\}$ $(1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J)$

or, equivalently,

over $\{p_{ij}\}$ $(1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J-1)$ and $\sum_{j=1}^{J} \alpha_{ij}$ $(1 \leqslant i \leqslant I)$

} in the case of model B

3.2 THE PROGRAMS

To find the maximum likelihood estimates, two programs were written in FORTRAN, one to estimate model A and the other to estimate model B. Both programs incorporate NAG routine EO4UAF (see Numerical Algorithms Group, 1981) to minimise

$$\sum_k \log|\Gamma_k^*| + \sum_k (\underset{\sim}{Y}_k^* - P^{*T}\underset{\sim}{X}_k)^T \Gamma_k^{*-1} (\underset{\sim}{Y}_k^* - P^{*T}\underset{\sim}{X}_k) \qquad (3.19)$$

(equivalent to maximising (3.18))

subject to the I inequality constraints

$$p_{ij} \geqslant 0 \ , \quad \sum_{j=1}^{J-1} p_{ij} \leqslant 1 \qquad (1 \leqslant i \leqslant I) \qquad (3.20)$$

Routine EO4UAF uses a sequential augmented Lagrangian method, the minimisation subproblems involved being solved by a quasi-Newton method. No derivatives are required - for discussion of this point see section 3.6.

The programs were written for the case $I = 4$, $J = 4$ and $K = \text{NCONST}$, a variable in the program to be set before running.

Accuracy - the parameter XTOL in routine EO4UAF was set at 0.0001, ensuring that the maximum likelihood estimates are accurate to at least three decimal places.

Details of the programs are given in Appendix 1 and Appendix 2.

3.3 GOODNESS OF FIT

The likelihood-ratio statistic is

$$\sum_k (\underset{\sim}{Y}_k{}^* - P^{*T}\underset{\sim}{X}_k)^T \Gamma_k{}^{*-1} (\underset{\sim}{Y}_k{}^* - P^{*T}\underset{\sim}{X}_k) \qquad (3.21)$$

on $\begin{cases} K(J-1) - I(J-1) \text{ degrees of freedom} & \text{(model A)} \\ K(J-1) - IJ \text{ degrees of freedom} & \text{(model B)}. \end{cases}$

The quantity (3.21) is calculated as part of (3.19) at each iteration of the maximisation routine and so is readily available for checking goodness of fit at any stage.

3.4 TESTING THE PROCEDURE FOR MODEL A

In all of this work the data used was extracted from the results of the two British General Elections in February and October 1974. In those elections there were 244 constituencies, from a total of 635, in which the Conservative, Labour and Liberal parties all contested both elections and no other parties were involved. For these 244 constituencies, then, we have $I = J = 4$ and the above programs can be applied.

Let $i = j = \begin{cases} 1 & \text{Conservative} \\ 2 & \text{Labour} \\ 3 & \text{Liberal} \\ 4 & \text{No vote} \end{cases}$

(i) Testing on simulated data

From the 244 constituencies above, 20 were selected at random. From the results of the February election in these constituencies, "results" for the second election were simulated using a random sample generated from a particular model A, with parameters

$$P = \begin{pmatrix} 0.64 & 0.01 & 0.16 & 0.19 \\ 0.01 & 0.80 & 0.06 & 0.13 \\ 0.22 & 0.03 & 0.32 & 0.43 \\ 0.02 & 0.05 & 0.02 & 0.91 \end{pmatrix} \qquad (3.22)$$

Details of the generation of the random sample are given in Appendix 4.

The results of the test were as follows: from starting values $\{p_{ij}\}$ as in (3.22), the maximisation algorithm coverged in 42 iterations to the matrix

$$P_{(42)} = \begin{pmatrix} 0.6405 & 0.0095 & 0.1569 & 0.1931 \\ 0.0072 & 0.7998 & 0.0632 & 0.1298 \\ 0.2196 & 0.0268 & 0.3310 & 0.4226 \\ 0.0235 & 0.0533 & 0.0088 & 0.9144 \end{pmatrix} \qquad (3.23)$$

with goodness-of-fit statistic 47.87 on 48 degrees of freedom.

(ii) Testing on real data

McCarthy and Ryan (1977) also analysed the results of the two 1974 elections. They found only 240 constituencies with the property described above; the discrepancy of 4 remains a mystery, but is possibly due to ambiguities in the categorisation of "independent Liberal" and other such candidates. One of the classifications used by McCarthy and Ryan to divide their 240 constituencies into smaller homogeneous groups was by Liberal strength; they defined Liberal strength for a constituency as the percentage of the electorate who voted Liberal at the February election. This enabled them to divide the 240 constituencies into three groups of 80, comprising "Low", "Medium" and "High" Liberal strength. As test data, we will use the results from the 80 constituencies (out of 244) with lowest Liberal strength; the Liberal strengths for this group range from 8.9% to 17.5% of the electorate.

The results of the test were as follows : from starting values

$$P_{(0)} = \begin{pmatrix} 0.90 & 0.05 & 0.03 & 0.02 \\ 0.10 & 0.80 & 0.05 & 0.05 \\ 0.03 & 0.02 & 0.90 & 0.05 \\ 0.02 & 0.02 & 0.02 & 0.94 \end{pmatrix}$$

the maximisation algorithm converged in 71 iterations to

$$P_{(71)} = \begin{pmatrix} 0.7904 & 0 & 0.1246 & 0.0850 \\ 0 & 0.8623 & 0.0253 & 0.1124 \\ 0.1677 & 0.2013 & 0.4616 & 0.1694 \\ 0.0120 & 0.0682 & 0 & 0.9198 \end{pmatrix} \qquad (3.24)$$

This matrix will be compared in Chapter 5 with the matrix obtained using a quadratic programming technique by McCarthy and Ryan (1977).

The goodness-of-fit statistic for this matrix is 70443.8 on 228 degrees of freedom, so model A is clearly not a good fit to these data.

3.5 TESTING THE PROCEDURE FOR MODEL B

(i) On simulated data

The simulated data used here were based on the 80 constituencies of lowest Liberal strength as described in section 3.4. In each of these constituencies the results of the first election were taken and combined with computer-generated random samples from $I(=4)$ particular compound multinomial distributions to give simulated results for the second election. Details of the computer generation are given in Appendix 5; the parameters supplied for the 4 compound multinomial distributions were

$$\alpha = (\alpha_{ij}) = \begin{pmatrix} 41 & 1 & 6 & 5 \\ 2 & 16 & 3 & 1 \\ 1 & 1 & 12 & 3 \\ 5 & 3 & 4 & 43 \end{pmatrix} \qquad (3.25)$$

In terms of p_{ij}'s and $\sum_j \alpha_{ij}$'s this is equivalent to

$$P^* = \begin{pmatrix} 0.774 & 0.019 & 0.113 \\ 0.091 & 0.727 & 0.136 \\ 0.059 & 0.059 & 0.706 \\ 0.091 & 0.055 & 0.073 \end{pmatrix}, \quad A = \begin{pmatrix} 53 \\ 22 \\ 17 \\ 55 \end{pmatrix} \qquad (3.26)$$

where $A_i = \sum_{j=1}^{J} \alpha_{ij} \qquad (i = 1,2,3,4)$

The starting values supplied to the maximisation algorithm were those in (3.26). After 90 iterations the algorithm had failed to converge to a maximum to the required accuracy and the program was stopped. The values reached after 90 iterations were

$$P^*_{(90)} = \begin{pmatrix} 0.801 & 0 & 0.120 \\ 0.089 & 0.733 & 0.140 \\ 0.029 & 0.101 & 0.706 \\ 0.056 & 0.078 & 0.035 \end{pmatrix}, \quad \underset{\sim}{A}_{(90)} = \begin{pmatrix} 58.6 \\ 23.4 \\ 29.3 \\ 17.7 \end{pmatrix} \qquad (3.27)$$

It was observed that over the last few iterations the values in the first two rows of P^* and $\underset{\sim}{A}$, i.e. the parameters for transitions from Conservative and Labour, had changed very little; their values after 90 iterations are also very close to those expected from (3.26). The third and fourth rows, however, were still changing quite a lot when the algorithm was stopped, and their values after 90 iterations are not as close to those we would have expected. These rows contain the parameters for transitions from the Liberals and Non-voters.

A likely explanation for this apparent indeterminacy in the third and fourth rows lies in the nature of the 80 constituencies under consideration here. The abstention rate is roughly constant in all the constituencies and in these particular 80 constituencies so is the Liberal strength, varying only between 8.9% and 17.5% of the electorate. This gives a very narrow base for estimation in these rows, so we might expect the parameter estimates to be ill-determined. The Conservative and Labour parties, however, each have a wide range of strengths represented in these 80 constituencies, so we would expect estimates of parameters for transitions from these parties to be somewhat more precise.

(ii) <u>On real data</u>

Here the actual results at both elections for the same 80 constituencies of low Liberal strength were used. The starting values supplied were

$$P^*_{(0)} = \begin{pmatrix} 0.905 & 0.005 & 0.084 \\ 0.001 & 0.935 & 0.001 \\ 0.001 & 0.131 & 0.610 \\ 0.001 & 0.001 & 0.001 \end{pmatrix}, \quad \underset{\sim}{A}_{(0)} = \begin{pmatrix} 100 \\ 100 \\ 100 \\ 100 \end{pmatrix}$$

The algorithm found maximum likelihood estimates after 125 iterations. Again it was noticed that the parameters in the last two rows of P* and A, i.e. those for transitions from the Liberals and Non-voters, were slow to be determined, causing the maximisation to require so many iterations. The final estimates were

$$P^*_{(125)} = \begin{pmatrix} 0.8755 & 0 & 0.0185 \\ 0 & 0.9358 & 0 \\ 0 & .1348 & 0.7114 \\ 0 & 0 & 0 \end{pmatrix}, \quad \underset{\sim}{A}_{(125)} = \begin{pmatrix} 26.2 \\ 97.0 \\ 25.0 \\ - \end{pmatrix} \qquad (3.28)$$

These estimates are analysed further in Chapter 5.

Note that the presence of a <u>zero</u> in any row of $P = (p_{ij})$ indicates an <u>improper</u> Dirichlet distribution. This causes no problems where there are at least two non-zero elements in a row, because of the following <u>theorem</u>:

If $(x_1,\ldots,x_r)$ is a random variable having the r-variate Dirichlet distribution $D(\nu_1,\nu_2,\ldots,\nu_r\,;\,\nu_{r+1})$ then the random variable $(z_1,z_2,\ldots,z_s)$ where

$$z_1 = x_1 + \ldots + x_{r_1}$$

$$z_2 = x_{r_1+1} + \ldots + x_{r_1+r_2}$$

$$\vdots$$

$$z_s = x_{r_1+r_2+\ldots+r_{s-1}+1} + \ldots + x_{r_1+r_2+\ldots+r_s}$$

and $\quad r_1 + r_2 + \ldots + r_2 \leqslant r$

has the s-variate Dirichlet distribution $D(\nu_{(1)},\ldots,\nu_{(s)}\,;\,\nu_{(s+1)})$

where

$$\nu_{(1)} = \nu_1 + \ldots + \nu_{r_1}$$

$$\vdots$$

$$\nu_{(s)} = \nu_{r_1+r_2+\ldots+r_{s-1}+1} + \ldots + \nu_{r_1+r_2+\ldots+r_s}$$

$$\nu_{(s+1)} = \nu_{r_1+r_2+\ldots+r_{s+1}} + \ldots + \nu_{r+1}$$

For proof of this theorem see, for example, Wilks (1962).

The implication of this theorem is that, as long as at least two of the α_{ij} in a particular row i are non-zero, any problem of improper random variables can be resolved by pooling elements. For example, in row 1 of (3.28) the first two elements 0.8755 and 0 can be added together. In row 4, however, there are three zero elements, the fourth element being $p_{44} = 1$. Here the improper distribution cannot be resolved by pooling; the result of this is that $\sum_{j=1}^{4} \alpha_{4j} = \alpha_{44} \equiv A_4$ is completely undetermined.

3.6 POSSIBLE ACCELERATION OF THE METHOD

It has been observed that the optimisation routine EO4UAF has taken between 42 and 125 iterations to find a maximum to the required accuracy. Whilst this is not a ridiculously large number of iterations, neither is it very quick.

These figures were achieved without first derivatives. It is commonly thought that if analytical first derivatives are supplied in an optimisation problem then computational efficiency is likely to be improved. We wish to maximise

$$L = -\frac{1}{2}\sum_k \log|\Gamma_k^*| - \frac{1}{2}\sum_k (Y_k^* - P^{*T}X_k)^T \Gamma_k^{*-1} (Y_k^* - P^{*T}X_k) \qquad (3.29)$$

When estimating model A, for example, we maximise L over $\{p_{ij} : 1 \leqslant i \leqslant I, 1 \leqslant j \leqslant J-1\}$. The matrix of partial derivatives in this case is given by

$$\frac{\partial L}{\partial P^*} = \Bigg(-\frac{1}{2}\sum_k \Big[\underset{\sim}{X}_k \underset{\sim}{1}^T \mathrm{Diag}(\Gamma_k{}^{*^{-1}}) - 2\,\mathrm{Diag}(\underset{\sim}{X}_k) P^*\Gamma_k{}^{*^{-1}} \Big]$$

$$- \frac{1}{2}\sum_k \Big[-2\underset{\sim}{X}_k (\underset{\sim}{Y}_k{}^* - P^{*^T}\underset{\sim}{X}_k)^T \Gamma_k{}^{*^{-1}} - \underset{\sim}{X}_k \left\{ \Gamma_k{}^{*^{-1}} (\underset{\sim}{Y}_k{}^* - P^{*^T}\underset{\sim}{X}_k) \right\}^{2^T} \qquad (3.30)$$

$$+ 2\,\mathrm{Diag}(\underset{\sim}{X}_k) P^*\Gamma_k{}^{*^{-1}} (\underset{\sim}{Y}_k{}^* - P^{*^T}\underset{\sim}{X})(\underset{\sim}{Y}_k{}^* - P^{*^T}\underset{\sim}{X})^T \Gamma_k{}^{*^{-1}} \Big] \Bigg)$$

where $1^T = (1,1,1,\ldots,1)_{J-1}$

and where $\mathrm{Diag}\ (\Gamma_k{}^{*^{-1}})$ means the diagonal matrix with the same diagonal entries as those in $\Gamma_k{}^{*^{-1}}$.

To calculate these derivatives at every iteration of an optimisation algorithm would require a considerable amount of computation; it is open to question, therefore, whether the inclusion of first derivatives in this case would lead to any significant increase in computational efficiency. This doubt, together with the amount of programming effort required to incorporate these derivatives into the estimation routines, prohibited any further investigation into acceleration by this method.

4. USE OF THE E-M ALGORITHM

An outline of the use of the E-M algorithm to estimate voter transition matrices was given at the end of Chapter 2. Here we give details of the way in which each step was constructed, and the results of testing the procedure on both simulated data and real data. For the moment we shall consider estimation only of model A, i.e. with fixed transition probabilities for all K constituencies.

4.1 INITIAL ESTIMATE OF Π

For reasons of computational efficiency, the initial estimate should clearly be as good a guess at the transition probabilities as possible. One necessary feature, however, is that the initial estimate of Π should not contain any zero elements; for once any Π_{ij} has become zero, this algorithm never changes its value again.

4.2 THE EXPECTATION (E-) STEP

This step must take the vectors $\underset{\sim}{X}_k$, $\underset{\sim}{Y}_k$ for each constituency k, and the present estimate of Π, and evaluate

$$t_{ijk} = E(T_{ijk} \mid \underset{\sim}{X}_k, \underset{\sim}{Y}_k, \Pi) \qquad \begin{pmatrix} 1 \leq i \leq I \\ 1 \leq j \leq J \\ 1 \leq k \leq K \end{pmatrix} \qquad (4.1)$$

Now the distribution of T_{ijk} conditional on $\underset{\sim}{X}_k, \underset{\sim}{Y}_k, \Pi$ is an extended hypergeometric distribution, defined by

$$h(t_{ijk} \mid \underset{\sim}{X}_k, \underset{\sim}{Y}_k, \Pi) = \frac{\{\exp(\sum_{i,j} t_{ijk}\lambda_{ij}) / \prod_{i,j} t_{ijk}!\}}{\sum_{t_{ijk}} \{\exp(\sum_{i,j} t_{ijk}\lambda_{ij}) / \prod_{i,j} t_{ijk}!\}} \qquad (4.2)$$

where

$$\lambda_{ij} = \log\ (\Pi_{ij}\Pi_{IJ} / \Pi_{iJ}\Pi_{Ij}) \qquad \begin{pmatrix} 1 \leq i \leq I \\ 1 \leq j \leq J \end{pmatrix} \qquad (4.3)$$

i.e. λ_{ij} is the logarithm of the cross-product ratio for the (i,j)th element of Π.

(See, for example, Plackett (1981))

Unfortunately, there are no simple explicit expressions for the moments of this distribution; Plackett (1981) states that "the exact expressions are inconvenient to use without good computational facilities, and for large values of n_{ij} [$\equiv T_{ijk}$ here] they eventually become useless." Since we can certainly expect large values of T_{ijk} in this problem, we will seek an approximation to the expectation. Following Plackett, we define (unique) quantities $\tau_{ijk}(\underset{\sim}{\lambda})$ to satisfy

$$\sum_{j=1}^{J} \tau_{ijk}(\underset{\sim}{\lambda}) = X_{ik} \quad , \quad \sum_{i=1}^{I} \tau_{ijk}(\underset{\sim}{\lambda}) = Y_{jk} \quad \text{(i.e. row and colum sums correct)} \tag{4.4}$$

and $$\log \{\tau_{ijk}(\underset{\sim}{\lambda})\tau_{IJk}(\underset{\sim}{\lambda})/\tau_{iJk}(\underset{\sim}{\lambda})\tau_{Ijk}(\underset{\sim}{\lambda}) = \lambda_{ij} \quad \forall i,j,k \tag{4.5}$$

(i.e. all cross-product ratios agree with those in Π)

Then the $\{T_{ijk}\}$ for constituency k can be shown to have an asymptotic multivariate normal distribution, with asymptotic means

$$E(T_{ijk} | \underset{\sim}{X}_k, \underset{\sim}{Y}, \underset{\sim}{\lambda}) \sim \tau_{ijk}(\underset{\sim}{\lambda}) \tag{4.6}$$

So we can return the values τ_{ijk} as the outcome of the E-step.

The problem now is that of finding quantities (τ_{ijk}) which satisfy (4.4) and 4.5); this is simplest done using the so-called iterative scaling procedure (ISP). This can be specified as follows: given, for constituency k, $\underset{\sim}{X}_k$, $\underset{\sim}{Y}_k$ and the latest estimate of Π,

(i) let $M_{ij} = X_{ik}\Pi_{ij}$ $\quad (1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J)$

(ii) scale the <u>columns</u> of $M = (M_{ij})$ to make column totals agree with the Y_{jk}, i.e. replace M_{ij} by $M_{ij} \times (Y_{jk} / \sum_{i=1}^{I} M_{ij})$

(iii) scale the <u>rows</u> of M to make row totals agree with the X_{ik}, i.e. replace M_{ij} by $M_{ij} \times (X_{ik} / \sum_{j=1}^{J} M_{ij})$

(iv) check for convergence of the sequence of matrices M; if not converged, return to (ii); if converged, return the values $\tau_{ijk} = M_{ij}$ $\quad (1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J)$.

It can be shown that this iterative procedure is guaranteed to converge, to any desired accuracy - see, for example, Bishop, Fienberg and Holland (1975).

Step (i) here ensures that the $\{M_{ij}\}$ have the correct cross-product ratios; steps (ii) and (iii) leave all cross-product ratios unchanged, and when the iteration has converged both row and column sums of **M** will be correct. Hence the final $\{M_{ij}\}$ will indeed satisfy (4.4) and (4.5).

4.3 THE MAXIMISATION (M-) STEP

In the situation of complete data, with all the transition counts T_{ijk} known, the likelihood function is

$$L(\Pi) = \prod_{k=1}^{K} \prod_{i=1}^{I} X_{ik}! \prod_{j=1}^{J} \frac{\Pi_{ij}^{T_{ijk}}}{T_{ijk}!} \quad (4.7)$$

$$\Rightarrow \log L(\Pi) = \sum_{k=1}^{K} \sum_{i=1}^{I} (\log(X_{ik}!) + \sum_{j=1}^{J} (T_{ijk} \log\Pi_{ij} - \log(T_{ijk}!))) \quad (4.8)$$

We wish to maximise $\log L$ subject to the constraints

$$\sum_{j=1}^{J} \Pi_{ij} = 1 \qquad (1 \leqslant i \leqslant I) \quad (4.9)$$

Putting in Lagrange multipliers λ_i $(1 \leqslant i \leqslant I)$, we need to solve

$$\sum_{k=1}^{K} \frac{T_{ijk}}{\Pi_{ij}} - \lambda_i = 0 \quad (4.10)$$

$$\Rightarrow \quad \Pi_{ij} = \sum_{k=1}^{K} T_{ijk}/\lambda_i \quad (4.11)$$

$$\Rightarrow \quad \lambda_i = \sum_{j=1}^{J} \sum_{k=1}^{K} T_{ijk}$$ to satisfy the constraints.

So the maximum likelihood estimates for the elements of Π are the obvious ones, i.e.

$$\Pi_{ij} = \sum_{k=1}^{K} T_{ijk} / \sum_{j=1}^{J} \sum_{k=1}^{K} T_{ijk} = \sum_{k=1}^{K} T_{ijk} / \sum_{k=1}^{K} X_{ik} \qquad 1 \leq i \leq I,\ 1 \leq j \leq J \tag{4.12}$$

We now replace T_{ijk} with the values τ_{ijk} obtained from the E-step, to get the new estimates

$$\hat{\Pi}_{ij} = \sum_{k=1}^{K} \tau_{ijk} / \sum_{k=1}^{K} X_{ik} \qquad 1 \leq i \leq I,\ 1 \leq j \leq J \tag{4.13}$$

This completes the M-step.

4.4 CONVERGENCE CRITERION

The test for convergence which was used in the initial testing of the E-M algorithm was as follows : denote successive iterates produced by the M-step as $\Pi_{(1)}, \Pi_{(2)}, \Pi_{(3)}, \ldots,$ and stop the algorithm at the n^{th} iteration if

$$d(\Pi_{(n)}, \Pi_{(n-1)}) < D$$

where $d(\Pi,\Pi') = \max_{i,j} |\Pi_{ij} - \Pi_{ij}'|$

and D is some threshold level to be fixed.

4.5 FOLLOWING PROGRESS : GOODNESS OF FIT

Writing $P^* = \begin{pmatrix} \Pi_{11} & \cdots & \Pi_{1,J-1} \\ \vdots & & \vdots \\ \Pi_{I1} & \cdots & \Pi_{I,J-1} \end{pmatrix}, \quad \underset{\sim}{Y}_k^{*T} = (Y_{1k}, Y_{2k}, \ldots, Y_{J-1,k})$

as in Chapter 3, we can evaluate the normal approximation to the log likelihood, which is proportional to

$$-\sum_k \log|\Gamma_k^*| - \sum_k (\underset{\sim}{Y}_k^* - P^{*T}\underset{\sim}{X}_k)^T \Gamma_k^{*-1} (\underset{\sim}{Y}_k^* - P^{*T}\underset{\sim}{X}_k) \tag{4.14}$$

and also the goodness-of-fit statistic

$$\sum_k (\underset{\sim}{Y}_k^* - P^{*T}\underset{\sim}{X}_k)^T \Gamma_k^{*-1} (\underset{\sim}{Y}_k^* - P^{*T}\underset{\sim}{X}_k) \tag{4.15}$$

These can be evaluated at each iteration as a check on how quickly the likelihood is increasing and the goodness-of-fit statistic is decreasing.

4.6 THE PROGRAM

A computer program was written in FORTRAN to perform the E-M algorithm in the case $I = 4$, $J = 4$ and $K = \text{NCONST}$, a variable in the program to be set before running. The program is given in Appendix 3.

4.7 TESTING THE PROCEDURE

(i) On simulated data

The algorithm was tested on the same simulated data from model A that were used in Chapter 3, i.e. on simulated data for 20 constituencies. The convergence threshold level D was set very low to allow a large number of iterations of the E-M algorithm; a maximum of 50 iterations was set in the program, however.

Recall that the data were simulated from the matrix

$$\Pi = \begin{pmatrix} 0.64 & 0.01 & 0.16 & 0.19 \\ 0.01 & 0.80 & 0.06 & 0.13 \\ 0.22 & 0.03 & 0.32 & 0.43 \\ 0.02 & 0.05 & 0.02 & 0.91 \end{pmatrix}$$

This matrix was supplied to the E-M algorithm as the initial estimate for Π.

The result of the test was that the algorithm did not satisfy the convergence criterion in the first 50 iterations, as expected since the convergence threshold level D was set very low at 10^{-6}. The 50th iteration was

$$\Pi^{(50)} = \begin{pmatrix} 0.6405 & 0.0098 & 0.1578 & 0.1918 \\ 0.0096 & 0.8003 & 0.0599 & 0.1302 \\ 0.2201 & 0.0286 & 0.3244 & 0.4270 \\ 0.0196 & 0.0500 & 0.0190 & 0.9114 \end{pmatrix}$$

It was noted that the normal approximation to the likelihood increased each iteration, and the goodness-of-fit statistic (4.15) decreased at each iteration, with a value after 50 iterations of 51.47; this is on 48 degrees of freedom and hence statistically non-significant.

However, the sequence $\Pi_{(r)}$ was clearly being very slow to converge. The behaviour of $\Pi_{11(r)}$ was studied more closely; the last eleven values were:

r	$\Pi_{11(r)}$	$\Pi_{11(r)}-\Pi_{11(r-1)}$	$\frac{\Pi_{11(r)}-\Pi_{11(r-1)}}{\Pi_{11(r-1)}-\Pi_{11(r-2)}}$ (estimates rate of convergence)
40	0.640430		
41	0.640438	0.000008	
42	0.640447	0.000009	$\approx$ 1.1
43	0.640455	0.000008	$\approx$ 1
44	0.640463	0.000008	$\approx$ 1
45	0.640471	0.000008	$\approx$ 1
46	0.640479	0.000008	$\approx$ 1
47	0.640487	0.000008	$\approx$ 1
48	0.640495	0.000008	$\approx$ 1
49	0.640503	0.000008	$\approx$ 1
50	0.640511	0.000008	$\approx$ 1

Table (4.16)

This behaviour suggests order 1 convergence at an extremely slow rate, very close to 1. The other elements of $\Pi_{(r)}$ were seen to behave in a very similar way.

(ii) <u>On real data</u>

The procedure was tested on the same 80 constituencies with low Liberal support that were used in Chapter 3. Once more the likelihood (4.14) increased and the goodness-of-fit statistic (4.15) decreased at each iteration, with a goodness-of-fit value of 71479.9 after 50 iterations (c.f. final value of 70443.8 from the maximum likelihood method of Chapter 3). The elements of $\Pi_{(50)}$ were quite close to the final values given by the method of Chapter 3, being in agreement to 2 decimal places in every case. Again, though, there was evidence of order 1 convergence; a table of the same type as (4.16) estimated the rate of convergence as 0.97, which is again extremely slow.

te: The subroutine designed to perfrom the iterative scaling procedure s tested separately. Convergence was found to be rapid (always in fewer an 15 iterations) even with very severe convergence criteria. Since this ocedure gives only the asymptotic expectations, however, a fairly mild nvergence criterion was eventually adopted : stop scaling at the n^{th} teration if

$$d(M_{(n)}, M_{(n-1)}) < 0.1$$

here again $d(M,M') = \max_{i,j} |m_{ij} - m'_{ij}|$

.8 ACCELERATION OF THE E-M ALGORITHM

The E-M algorithm appears to give order 1 convergence at a very slow te, greater than 0.95. This suggests the use of an accelerator such as tken's Δ^2 method:

heorem (Aitken's Δ^2 method)

Let $\{x_n\}$ be any sequence converging to the limit s such that the uantities $d_n = x_n - s$ satisfy

$$d_n \neq 0 \quad , \quad d_{n+1} = (A + \varepsilon_n) d_n \tag{4.17}$$

here A is a constant, $|A| < 1$ and $\varepsilon_n \to 0$ as $n \to \infty$. Then the sequence $\{x_n'\}$ defined by

$$x_n' = x_n - \frac{(x_{n-1} - x_n)^2}{x_{n+2} - 2x_{n+1} + x_n} \tag{4.18}$$

is defined for n sufficiently large and converges to s faster than the sequence $\{x_n\}$ in the sense that

$$\frac{x_n' - s}{x_n - s} \to 0 \quad \text{as} \quad n \to \infty \tag{4.19}$$

(For proof of this theorem see, for example, Henrici (1964)).

Aitken's Δ^2 method was tried on the sequence $\{\Pi_{(r)}\}$ generated from the simulated data, applying (4.18) to each of the sequences $\{\Pi_{ij(r)}\}$

separately $(1 \leq i \leq 4,\ 1 \leq j \leq 4)$. The attempt was unsuccessful because n = 50 was clearly not the "n sufficiently large" of the above theorem. Unfortunately time did not permit further investigation of this method.

2.9 APPLICATION OF THE E-M ALGORITHM TO THE ESTIMATION OF MODEL B

The problem here lies with the M- step. The E- step is much the same as for model A, but using the $\{p_{ij}\}$ of (2.10) in place of the $\{\Pi_{ij}\}$ of model A. However, obtaining maximum likelihood estimates for the parameters of the compound multinomial distribution from the complete data is not as simple as it was for model A. Mosimann (1962) gives a number of possible estimators for the parameters of the compound multinomial distribution. However it is also pointed out that none of these is the maximum likelihood estimator, which can only be obtained by numerical maximisation of the likelihood function.

In principle, therefore, it would be possible to apply the E-M algorithm to the problem of estimating the parameters of model B; in practice, however, it is likely to be computationally very slow and therefore of little value.

5. FURTHER ANALYSIS OF THE 1974 DATA

.1 THE DATA

The data to be analysed are the results of the February and October itish General Elections of 1974 in the 244 constituencies where there s a straight fight between the Conservative, Labour and Liberal parties both elections. The aggregate results over these 244 constituencies re as follows:

Table (5.1)

		Total Votes (millions)	% of Electorate	Number of seats
BRUARY	CON	5.370	33.78	166
	LAB	4.153	26.13	72
	LIB	3.272	20.58	6
	NV	3.101	19.51	-
TOBER	CON	4.949	30.89	159
	LAB	4.240	26.46	81
	LIB	2.657	16.58	4
	NV	4.177	26.07	-

Here "NV" indicates Non-voters.

We can see from these aggregate figures that the sample of 244 nstituencies rather over-represents those held by the Conservatives. Carthy and Ryan (1977) point out that minor party intervention is more evalent in urban areas, which return predominantly Labour members.

The above table (5.1) shows that turnout fell between the two elections, that the Conservative and Liberal parties lost a large number of votes ile Labour gained a few.

.2 ANALYSIS OF THE TRANSITION PROBABILITIES

The same data were analysed by McCarthy and Ryan (1977) although, as as pointed out in Section 3.4, they found only 240 constituencies with traight CON-LAB-LIB contests. They estimated voter transition matrices sing restricted least suqares, without weights, solved by a quadratic rogramming method. For purposes of comparison we will present the cCarthy and Ryan (M-R) estimates along with those found for models A nd B using the (restricted) maximum likelihood estimators described in hapter 3.

The estimates are given in table (5.2).

Table (5.2)

FEBRUARY		OCTOBER CON	OCTOBER LAB	OCTOBER LIB	OCTOBER NV	
	CON	0.897	0.025	0	0.077	(M - R)
		0.805	0	0.081	0.114	(Model A)
		0.897	0	0.002	0.101	(Model B)
	LAB	0	0.932	0	0.069	
		0	0.896	0	0.104	
		0	0.953	0	0.047	
	LIB	0.049	0.069	0.816	0.066	
		0.152	0.097	0.653	0.098	
		0.015	0.080	0.790	0.115	
	NV	0	0	0	1	
		0.022	0.058	0.017	0.903	
		0	0	0	1	

Points to note in comparing the three sets of estimates are:

(i) the very close agreement between the estimates from model B and those of McCarthy and Ryan;

(ii) the model A estimates are, on the whole, lower on the leading diagonal than those given by the other two methods, and higher off the leading diagonal;

(iii) all three methods of obtaining estimates give zero values for some parameters. This remains something of a problem; these values cannot be true, for it is unconceivable that, for example, no-one who voted Labour in February voted Liberal in October.

The general picture given by all three sets of estimates, however, is e same. Approximately equal percentages of the three main parties moved the Non-voting group. Labour supporters remained loyal (the estimates tually imply that no Labour supporters changed their vote), with the berals doing most vote-changing.

We will now consider the transition matrix for the group of 80 nstituencies with lowest Liberal strength, as defined in Section 3.4.

Table (5.3)

Estimated transition probabilities for the 80 constituencies with lowest Liberal strength

FEBRUARY	OCTOBER: CON	LAB	LIB	NV	
CON	0.911	0.005	0.084	0	(M-R)
	0.790	0	0.125	0.085	(Model A)
	0.875	0	0.019	0.106	(Model B)
LAB	0	0.936	0	0.064	
	0	0.862	0.025	0.113	
	0	0.936	0	0.064	
LIB	0	0.131	0.612	0.257	
	0.168	0.201	0.462	0.169	
	0	0.135	0.711	0.154	
NV	0	0	0	1	
	0.012	0.068	0	0.920	
	0	0	0	1	

ain we see that the estimates from model B are very close to those rthy and Ryan and that model A estimates are lower on the leading al and higher elsewhere.

he most striking difference between tables (5.2) and (5.3) is that n Liberal loyalty rates. The defection rate from the Libe~~ary~~ral Party ch higher in the constituencies where Liberal support was low than 244 constituencies as a whole - a feature which is not at all sing. The pattern of Liberal defection was also different; in tuencies with low Liberal strength a defecting Liberal was less y to move to the Conservatives than in the sample as a whole.

NTERPRETATION OF THE ESTIMATED SUMS $\sum_j \alpha_{ij}$ FOR MODEL B

Table (5.4)

Estimated sums $\sum_j \alpha_{ij}$ for model B

	For the full sample of 244 constituencies	For the 80 constituencies with lowest Liberal strength
(CON)	30.0	26.2
(LAB)	86.0	97.0
(LIB)	19.6	25.0
(NV)	-	-

The quantity $\sum_{j=1}^{4} \alpha_{ij}$ gives an indication of the amount of variability the transition probabilities from party i between constituencies. The ller the value of $\sum_j \alpha_{ij}$, the more variability there is.

The estimates in table (5.4) indicate that there is little variability tween constituencies in the behaviour of Labour voters, but more bility in the behaviour of Conservative and Liberal voters. It is y that one of the major sources of variability between constituencies actical voting" whereby, in a marginal contest between two of the parties, supporters of the third party vote not for their own urite but for the "lesser of two evils". The above estimates, then, t be taken as an indication that Conservatives and Liberals indulge actical voting to a greater extent than do Labour voters.

The actual variances of the probabilities (Π_{ijk}) may be calculated. If $(x_1,\ldots,x_r)$ has a Dirichlet distribution $D(\nu_1,\ldots,\nu_r;\nu_{r+1})$

$$\text{var}\,(x_i) = \frac{\nu_i(\nu_1+\ldots+\nu_{r+1}-\nu_i)}{(\nu_1+\ldots+\nu_{r+1})^2(\nu_1+\ldots+\nu_{r+1}+1)} \qquad (5.5)$$

e have

$$\text{var}\,(\Pi_{ijk}) = \frac{p_{ij}(1-p_{ij})}{\sum_{j=1}^{J}\alpha_{ij}+1} \qquad (5.6)$$

So, for example, for transitions from Labour we have

$$\text{var}\,(\Pi_{ijk}) \quad \frac{0.9 \times 0.1}{91} = 0.001$$

a standard deviation of the order of 0.03, and for transitions from Conservatives var (Π_{ijk}) $\frac{0.9 \times 0.1}{31} = 0.003$, i.e. a standard ation of the order of 0.05.

Covariances between the parameters may similarly be estimated using formula

$$\text{cov}\,(\Pi_{ijk}, \Pi_{ijk}) = -\frac{p_{ij}\,p_{ij'}}{\sum_{j=1}^{J}\alpha_{ij}+1} \qquad (j \neq j') \qquad (5.7)$$

A final point : we might have expected that the smaller group of 80 ituencies, having a certain political homogeneity, would have exhibited ariability in the transition probabilities. So we would have expected $\Sigma\,\alpha_{ij}$ to have larger values for the smaller group. It is a little sing, then, that $\sum_j \alpha_{ij}$, the value for transitions from the Conservatives ally estimated lower (at 26.2) in the smaller group than it is full sample (at 30.0). This feature could, of course, be explained pling variation; investigation of the sampling properties of the tor used would be useful to check this.

ANALYSIS OF RESIDUALS

The goodness-of-fit statistics for the two models on the two groups onstituencies were as follows:

Table (5.8)

Goodness-of-fit statistics

	Full sample of 244 constituencies	80 constituencies of low Liberal strength
l A	258117 (on 720 d.f.)	70443.8 (on 228 d.f.)
l B	732.379 (on 716 d.f.)	240.114 (on 224 d.f.)

Model B, then, gives a good fit to the data whereas model A clearly s not.

The goodness-of-fit statistic used is

$$\sum_{k=1}^{K} (\underset{\sim}{Y}_k{}^* - P^{*T}\underset{\sim}{X}_k)^T \Gamma_k{}^{*-1} (\underset{\sim}{Y}_k{}^* - P^{*T}\underset{\sim}{X}_k) \tag{5.9}$$

This may be decomposed to give residual "distances" for each nstituency,

$$D_k = (\underset{\sim}{Y}_k{}^* - P^{*T}\underset{\sim}{X}_k)^T \Gamma_k{}^{*-1} (\underset{\sim}{Y}_k{}^* - P^{*T}\underset{\sim}{X}_K) \tag{5.10}$$

We will examine these distances for model B fitted to the smaller oup of 80 constituencies. The goodness-of-fit statistic is, in theory, square random variable on 224 degrees of freedom; the observed of 240.114 is certainly not inconsistent with this. The quantities hen, might be expected to have almost a chi-quare distribution on rees of freedom. The observed distribution was as follows:

Table (5.11)

Observed distribution of the residual "distances"

Interval	Count
0-1	18
1-2	20
2-3	17
3-4	11
4-5	2
5-6	2
6-7	2
7-8	1
8-9	1
9-10	2
10+	4

is not inconsistent with a chi-square distribution on 3 degrees of dom.

It was felt that plotting these residual distances against constituency-ated variables might show up any pattern in the departures from the fitted l. The distances are plotted against size of constituency (Plot (5.12)) against the Conservative share of the two-party (i.e. CON+LAB) vote t (5.13)); one might also consider variables more directly related the social composition of the constituencies.

The plot of distances against size of constituency shows a tendency the smaller constituencies to be further away from the fitted model. re is no obvious explanation for this feature. It is, however, highly that size of constituency is highly correlated with other variables whether the constituency is urban or rural, and it may be that these "hidden" variables would explain departures from the fitted

t has been suggested in the past, for example by Miller (1972), ransition rates might be dependent on the partisanship of the tuency. As a measure of partisanship we take the Conservative of the two-party vote at the first election. Plot (5.13) shows ar relation between the residual distances and our measure of anship. An alternative to the distances of (5.10) is to consider tual residuals from the fitted model. For example, we define the rvative residuals" as

nservative residual = observed Conservative vote in October

- fitted Conservative vote in October

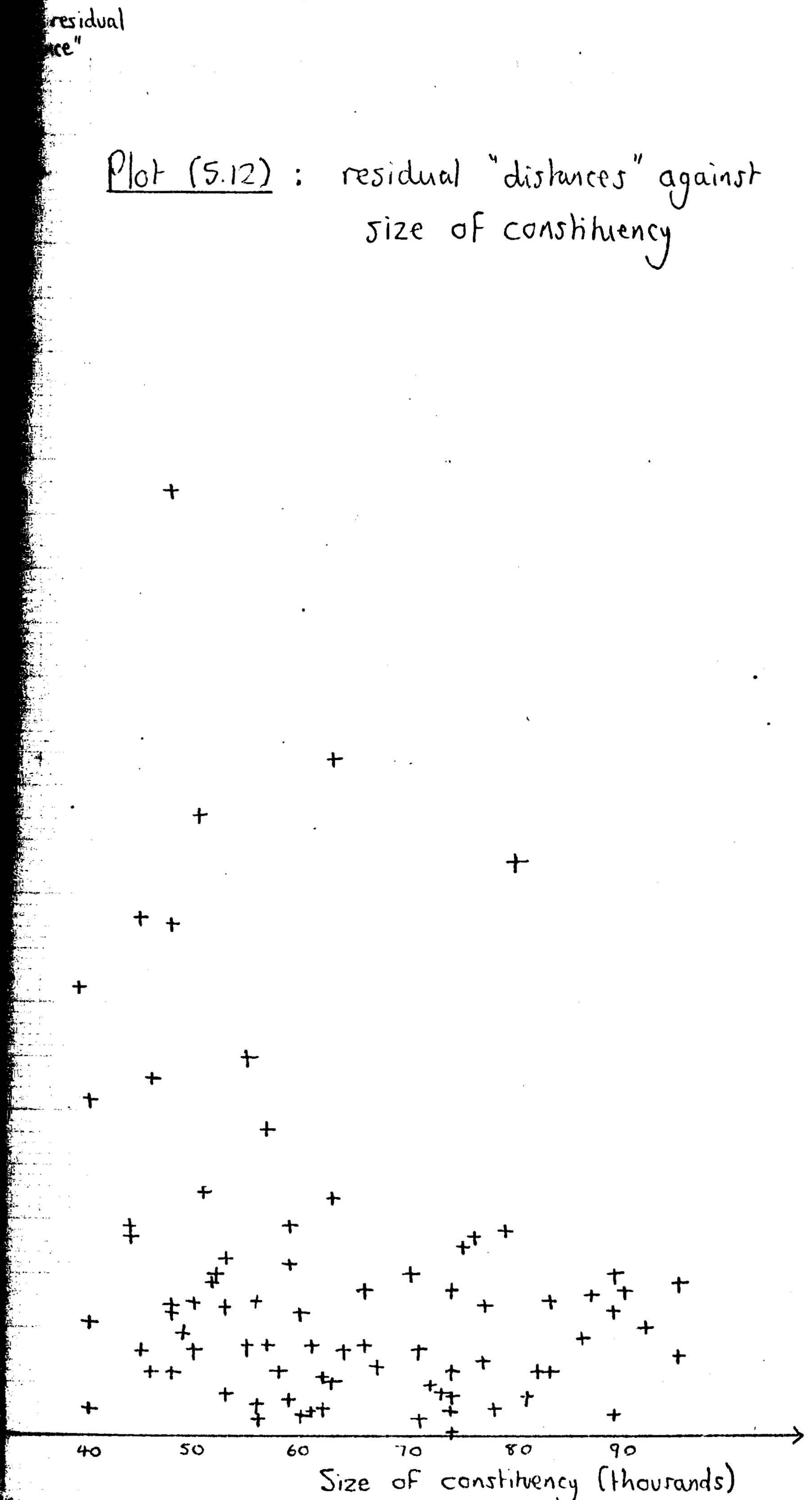
the fitted results are calculated from the February results and the ated probabilities (p_{ij}). This gives an indication of the direction parture from the model as well as its size. The Conservative residuals plotted against Conservative share of the two-party vote in Plot (5.14) ng a clear relationship, with the Conservative vote being over-ated where the Conservatives are weak and under-estimated where they trong. This suggests that a possible improvement to the model might o allow the parameters (α_{ij}) to vary with partisanship. The relation- could alternatively be investigated by classifying the constituencies groups according to partisanship and comparing the estimated (α_{ij}) different groups.

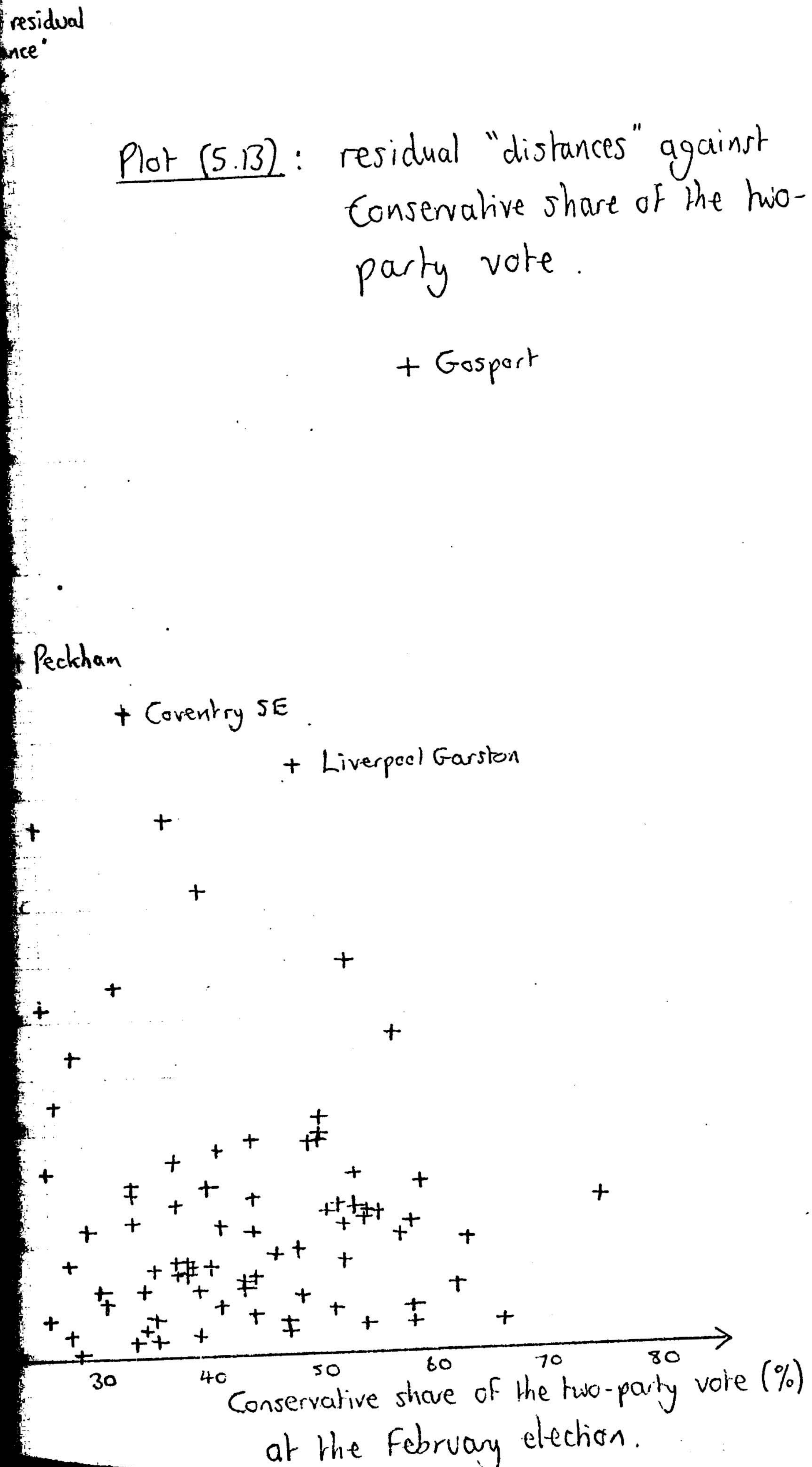
A note about large residuals : the largest residual distances in this sample of 80 constituencies with low Liberal strength were as follows:

Constituency	Residual distance, D_k
Gosport	17.54
Peckham	12.56
Coventry S E	11.56
Liverpool Garston	10.57

It would be of interest to discover whether there were any special umstances in these constituencies which might account for departure the model - there are none which are obvious just from the election lts.

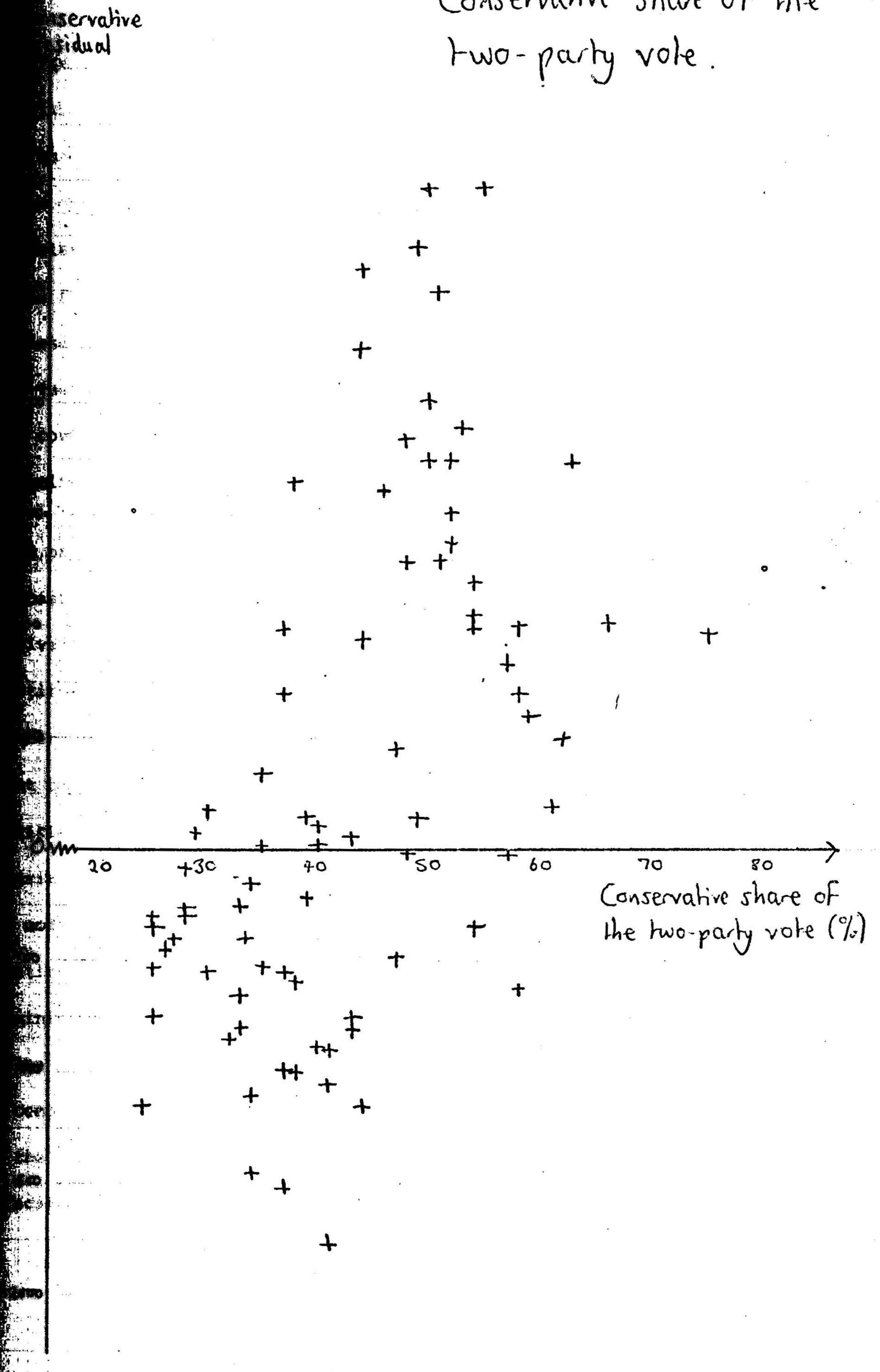
Plot (5.12) : residual "distances" against size of constituency





Plot (5.13): residual "distances" against Conservative share of the two-party vote.

Plot (5 14) : "Conservative residuals" against Conservative share of the two-party vote.



CONCLUDING REMARKS : POSSIBLE FURTHER WORK

he "random effects" model, model B, allowing some variation in tion probabilities between constituencies, was found to fit election ch better than the model postulating equal transition probabilities group of constituencies. The parameters of the underlying let distribution in the random effects model give not only average for transition probabilities but also the amount of variability se probabilities between constituencies. It would be useful to ust how precisely the parameters of this model can be estimated; ould be done either by evaluating the information matrix or using lation experiment.

---

problem not considered in this project is that of "election night sting", where the aim is to forecast the outcome of all constituencies ved in the election given the results of only the first few tuencies to have their votes counted. It is not immediately clear er the random effects model's "average" transition probabilities to do better or worse on the forecasting problem than the probties estimated for the ill-fitting fixed effects model. An interesting iment, then, would be to estimate probabilities, using each of the odels, from the results at two elections of a small number of tituencies, and then apply these estimates to the results of the ining constituencies at the first election to forecast the results he second election. The forecast results from each model could then ompared with the actual, observed results, giving at least some idea the relative merits of "average" versus "fixed" probabilities for ecasting.

-M algorithm, although conceptually and computationally very e for estimating the transition probabilities in the fixed effects l, proved to have order 1 convergence at a very slow rate. One ble cause of the slow convergence may be the approximation taken e "E-" step, but it has been reported by other authors that the algorithm can be very slow in certain cases - see the discussion e paper by Dempster, Laird and Rubin (1977). The possibility of erating the algorithm using Aitken's Δ^2-method or some other d was discussed briefly in Section 4.8, and it might be worth- to pursue this further.

---

e two main points to arise from the analysis of the two General ions of 1974 were

i) the differences in the behaviour of Liberals between constituencies of low Liberal strength and those of high Liberal strength

ii) the low degree of variability between constituencies found in the transition rates from the Labour Party, possibly suggesting that tactical voting is less common among Labour voters than among Conservatives and Liberals.

alysis of the residuals from the random effects model shows a onship between departure from the model and the partisanship of a tuency. The next stage in modelling, then, might be to allow the ters (α_{ij}) of the underlying Dirichlet distributions to vary with anship.

REFERENCES

Y.M.M., FIENBERG, S.E. and HOLLAND, P.W. (1975). Discrete ltivariate Analysis : Theory and Practice. Cambridge, Mass. : T Press.

R, A.P., LAIRD, N.M. and RUBIN, D.B. (1977). Maximum Likelihood rom Incomplete Data via the EM Algorithm. J.R. Statist.Soc. B, 1 - 38.

, A.G. (1969). An Approach to the Analysis of Electoral Swing. .R. Statist. Soc. A, 132, 68 - 79.

, P. (1964). Elements of Numerical Analysis. London : Wiley.

G. and MEETER, D. (1969). Building Voter Transition Models from ggregate Data. Midwest J. Polit. Sci., 13, 545 - 566.

N, N.J. and KOTZ, S. (1969). Discrete Distributions. Boston : oughton Mifflin.

HY, C. and RYAN, T.M. (1977). Estimates of Voter Transition robabilities from the British General Elections of 1974. J.Royal tatist.Soc. A, 140, 78- 85.

, W. (1972). Measures of Electoral Change using Aggregate Data. J.R.Statist. Soc. A, 135, 122 - 142.

NN, J.E. (1962). On the compound multinomial distribution, the ultivariate β-distribution and correlations among proportions. Biometrika, 49, 65 - 82.

ICAL ALGORITHMS GROUP (1981). NAG Library Manual Mk.8.

ETT, R.L. (1981). The Analysis of Categorical Data (Second Edition) London : Griffin.

S, S.S.(1962). Mathematical Statistics. London : Wiley.