The Brown and Payne model of voter transition revisited

Antonio Forcina and Giovanni M. Marchetti

Abstract We attempt a critical assessment of the assumptions, in terms of voting behavior, underlying the Goodman (1953) and the Brown and Payne (1986) models of voting transitions. We argue that the first model is only a slightly simpler version of the second which, however, is fitted in a rather inefficient way. We also provide a critical assessment of the approach inspired by King et al (1999) which has become popular among Sociologists and Political scientists.

Key words: Voters transitions; Ecological correlation; Over-dispersion.

1 Introduction

Estimating transitions of voters between two adjacent elections is one of extracting information on the association of a two way contingency table from its margins. As shown by Plackett (1977), in the 2×2 case, the information on the odds ratio provided by the margins of a single table is of a rather inconclusive nature. Goodman (1953) provided a formal statistical model which indicates that, under the assumption that a set of local units share the same pattern of transitions, this can be consistently estimated from the data. This result relies on the assumption that a set of tables, for which only the margins are observed, are determined by the same probabilistic model.

In the sociological literature the problem is seen as one of inferring individual behaviour from aggregate data and is known as *Ecological Inference*. In a famous

A. Forcina
Dipartimento di Economia, Finanza e Statistica, via Pascoli 10, 06100 Perugia, Italy; e-mail: forcina@stat.unipg.it

G. M. Marchetti
Dipartimento Statistico, Viale Morgagni, 59, 50134 Firenze, Italy; e-mail: giovanni.marchetti@ds.unifi.it

paper Robinson (1950) proved that the true underlying association at the individual level and that emerging from aggregate data may even have a different direction, a result know as the Ecological fallacy. As shown in Wakefield (2004, pag. 10), this an instance of the Simpson paradox and may arise when each local unit exhibit an association structure which is strongly correlated with the row marginal. Though this possibility is ruled out in the Goodman model, the model is still looked at with some suspicion (see, for instance, Anastasi et al, 1989), an attitude which has some justification because, due to its appealing simplicity, Goodman's model is sometimes used without attention to its underlying assumptions.

Brown and Payne (1986) proposed a model which, as we argue below, may be seen as an extension of Goodman's in an attempt to make its assumptions a little more realistic. This model has not become popular, perhaps because its estimation procedure is substantially more complex than the linear regression required for Goodman's model. Forcina and Marchetti (1989) reformulated the Brown and Payne model as a multivariate generalized linear model and made available a more efficient software.

An approach popular among Sociologists and Political scientists is due to King, Rosen and Tanner (1999, 2004) and is based on a hierarchical bayesian model. In section 3 we discuss this approach and some related methods. The Bayesian approach proposed by Bernardo (2001) is apparently based on assumptions similar to those of the Brown and Payne model, though the description of the model that he provides is not specified in sufficient detail.

The assumptions underlying Goodman's model and a modified version of the Brown and Payne model (FMBP) are discussed in sections 2 and 3. Concluding remarks are proposed in section 4. FMBP has been used to analyze voting transitions for most recent elections held in Umbria (Central Italy), see Bracalente, Ferracuti and Forcina (2006); reports appeared also on the local media.

2 The Goodman Model

Let n_u, $u = 1,\ldots,s$ denote the vector containing the number of voters in local unit u at election 1 $e1$ and y_u be the corresponding vector at election 2 $e2$. Suppose that the voting behavior of voters of party i $(i = 1,\ldots,I)$ at $(e2)$ satisfies the following assumptions:

1. the probability that a voter of party i at $e1$ chooses party j $(j = 1,\ldots,J)$ at $e2$ does not depend on the local unit u and is equal to $P(Y = j \mid X = i)$, where X,Y are the options selected at $e1$ and $e2$ respectively ;
2. voters decide independently of one another.

Let y_{iu} denote the vector containing the frequency distribution at $e2$ of voters in unit u who voted party i at $e1$; the above assumptions imply that y_{iu} is distributed as Multinomial(n_{iu}, p_i), where

$$p_i = (P(Y = 1 \mid X = i), \ldots, P(Y = J \mid X = i))'.$$

The y'_{iu}, $i = 1, \ldots, I$, may be seen as the rows of a frequency table which could be constructed if the choices of each voter at the two elections was known; in reality, only the row and columns totals can be observed. However, because $y_u = \sum_i y_{iu}$, simple algebra shows that the expectation of the vector of observed proportions $y_u/(1'n_u)$, being the sum of I multinomial random variables, is a linear function of the vectors of transition probabilities $p_1, \ldots, p_I$. Thus the transition probabilities could be estimated by multivariate linear regression.

However, the basic assumptions for optimality of ordinary least squares are violated in two directions:

1. the variance of $y_u/(1'n_u)$, the vector of observations in each local unit, equals the variance of a mixture of multinomial variables, thus it is not constant and depends on the unknown transition probabilities;
2. observations within the same local unit are not independent.

Though these violations affect only the efficiency of the estimates, when estimates are adjusted to force values to lie between 0 and 1, consistency of the estimates is also affected.

2.1 The Brown and Payne model revisited

Let us consider how realistic are the assumptions on which the multinomial model is based. It seems reasonable to believe that voters may affect each other within small circles; this may be due to personal interactions and to the fact that voters within the same local unit who selected the same party at $e1$ may be affected by common local peculiarities which may be difficult to detect and are naturally treated as random. In both case the multinomial assumption of independence would be violated and a different variance function would be adequate.

The BP model was motivated by the need to take this into account and, at the same time, avoid the inconveniences due to the method of estimation used to fit the Goodman model. The main features of the model are the following:

1. the vectors of transition probabilities are allowed to differ across local units as in a random effect model, more precisely

   $$p_{i|u} \sim Dirichelet(\pi_i, \psi_i)$$

   where now $p_{i|u}$ denotes the vector of transition probabilities from party i within local unit u; these probabilities are assumed to fluctuate around the overall average π_i with a covariance matrix given by $\psi_i[\mathrm{diag}(\pi_i) - \pi_i\pi_i']$; direct calculations show that

   $$\mathrm{Var}(y_{iu}) = n_{iu}[1 + \psi_i(n_{iu} - 1)][\mathrm{diag}(\pi_i) - \pi_i\pi_i'], \tag{1}$$

   which may be interpreted as the variance of an overdispersed multinomial;

2. maximum likelihood rather than least square estimates are computed, so the variance structure is taken into account;
3. transition probabilities π_i always lie between 0 and 1 because parameters of interest are defined by a multivariate logit transformation;
4. because the likelihood for a sum of overdispersed multinomial variables is almost untractable, a central limit approximation is used.

The problem with this approach is that the expression for $\mathrm{Var}(y_{iu})$, as can be seen from (1), is quadratic in the sample size; this makes the application of the central limit problematic. As an alternative, we propose a model of overdispersion where $\mathrm{Var}(y_{iu})$ is linear in n_{iu} and where the over dispersion parameter θ_i, as in the Brown and Payne (1986) model, is specific for each party. In the Appendix we show that this variance function is an approximation of the true variance under the following assumptions:

- the voters of party i in local unit u are composed of C_{iu} clusters of size n_{iuc} and their behaviour at $e2$ is determined by a vector of transition probabilities p_{iuc} which is sampled from a Dirichelet$(\theta_i, \pi_i$;
- the n_{iuc}, $c = 1, \ldots, C_{iu}$ are distributed as a multinomial with total n_{iu} and cell probability equal to $1/C_{iu}$, that is clusters tend to be of the same size;
- as the n_{iu} increase, n_{iu}/C_{iu} converges to a constant.

3 Recent alternative approaches

The hierarchical Bayesian model developed by King, Rosen and Tanner (1999) tries to exploit the fact that the frequency distribution of voters at two different elections in a given local unit determine a Frechet class of possible tables of voting transitions consistent with the given margins. Within this class the transition frequencies vary within well defined bounds and are linearly related. The model allows the vectors of transition probabilities to be specific for each local unit and relies on the crucial assumption that, conditionally on these transition probabilities, y_u is distributed as an overdispersed multinomial and not as a mixture of overdispersed multinomials as in the Brown and Payne model. This assumption, which simplifies computations, is equivalent to assume that all voters in local unit u are homogeneous with a common vector of transition probabilities equal to the weighted average of the transitions within each subgroup, an assumption which, we believe, is rather unrealistic.

The sets of bounds and linear relations between the transition probabilities within each local unit, as determined by the margins, are the basis of the approach developed by De Sio (2008). The idea is to compare the range of transition probabilities allowed by within each local unit with those allowed by the table obtained by marginalizing over local units. Is is assumed that the transition probabilities for a single table fluctuate around a common value and a search algorithm based on least squares is proposed to find a unique estimates with is least discrepant relative to the allowed range in each local unit and in the overall table. The limitation of this approach seems to be in a lack of a proper statistical model for the random fluctuations which justifies the optimization used in the algorithm.

4 An application

The approach described in Section 3 was applied to estimate transitions between the elections for the European Parliament and the one for the local administration held in the borough of Perugia (PG, central Italy) in June 2009. PG has slightly more than 126 thousands voters and is divided into 159 polling stations; however 4 of these were removed because they were located inside hospitals or the local prison. Though voters should approximately be the same for the two elections, this is not exactly true as shown by Fig. 1: the two stations with a relative difference larger than 6% were removed. Fig. 2 displays the Mahalanobis distance between the results in

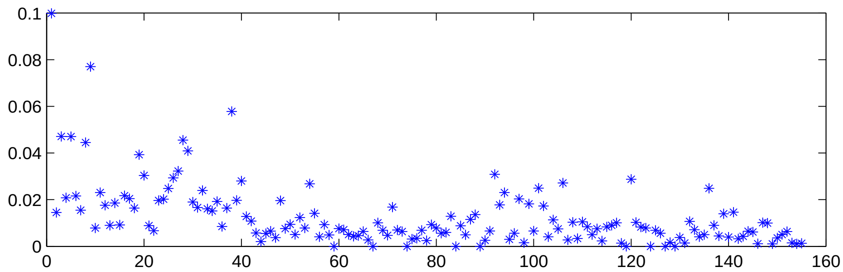

Fig. 1 Amount of relative difference between voters at European and Local election in each polling station

the local election and those predicted by the model as estimated in a preliminary analysis. The two polling stations with a discrepancy greater than 50 were removed from the final analysis.

Estimated transition probabilities together with standard errors computed from the expected information matrix and a Delta method are displayed in the table below where, for conciseness, the original transition matrix based on 10 rows and 12 columns has been condensed.

The PD, whose coalition won the local election but in the European election got less votes than the PdL, seems to have a high degree of fidelity. A surprising result is the large proportion of voters who, having supported one of the right wing party in the European election, seems to move to one of the left wing parties in the local election, a phenomenon which, given the local context, is considered plausible; note however that transitions from OR have very large standard errors. Finally note that, though some parties lost voters who abstained or gave a blank ballot, nobody who had abstained in the European election seems to have voted for the local administration.

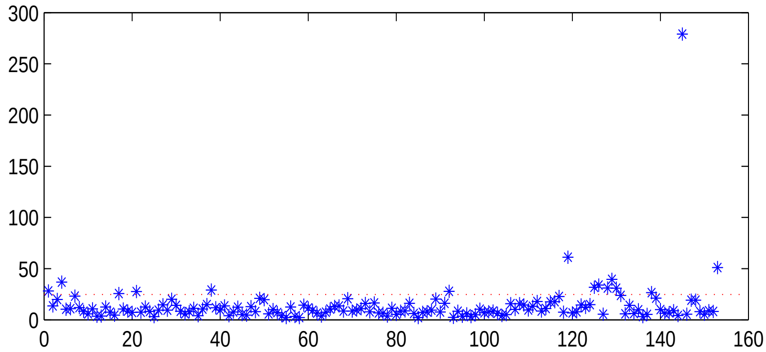

Fig. 2 Mahalanobis distance between observed and predicted electoral results by polling stations and 99% limit

Table 1 Estimated transition out of 100 voters and standard errors from the election for the European Parliament (row) to Local administration (column)in Perugia

Party	PD	se	OL	se	LL	se	UDC	se	PdL	se	OR	se	LR	se	NV	
PD	97.2	1.7	0.9	0.6	0.1	0.1	0.2	0.2	1.1	0.7	0.1	0.1	0.4	0.3	0.0	0.5
OL	0.5	1.0	67.3	3.0	14.6	1.8	0.0	1.4	3.9	1.9	1.9	1.1	5.1	2.2	6.7	2.7
UDC	0.0	0.1	15.3	6.0	0.0	0.1	68.1	5.2	10.0	5.5	0.0	0.0	6.6	4.2	0.0	2.4
PdL	0.0	0.5	5.5	2.2	1.4	1.0	1.7	1.1	71.6	2.0	0.5	0.8	11.6	1.8	7.8	2.1
OR	0.4	10.3	41.9	10.1	0.0	0.4	0.0	1.3	4.3	7.7	43.5	4.0	9.7	8.4	0.1	6.4
NV	0.1	1.6	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	99.9	2.1

PD=Democratic Party, OL=other party on the left, LL=local parties on the left, UDC=Union of Center Democrats, PdL=Party of the freedoms, OR=other parties on the right, LR=local parties on the right, NV=non voters.

5 Concluding remarks

Though the version of the Brown and Payne model presented above could still be improved by considering possible alternative models of overdispersion or by allowing the user to input subjective prior knowledge, it is superior to the Goodman's model fitted by linear regression. Relative to the hierarchical Bayesian model of King (1999), our model seems to be based on specific assumptions concerning voting behaviour and does not attempts to provide a general solution to the so called "Ecological Inference". As such, its use is recommended only within areas of limited dimension, like cities of medium size. The reason is that the assumption of a dominating pattern of voting transition would be unrealistic in very large metropolitan areas and even less if applied to a whole country. It is also important that local units, like the Italian polling stations, are reasonably small as the amount of information provided by aggregate data, obviously, decreases when local units of smaller size are clumped together.

In addition, the quality of the data, together with the scope of the application, are a crucial issue. The quality of the data requires that the voters within each local unit are, at least approximately, the same, a requirement which is rather problematic, at least in Italy, because boundaries of local units keep changing from time to time. This could be accommodated simply by merging units which have been redesigned by internal shifts. Accurate information must also be collected to spot special local units like hospitals whose voters may be expected to be completely different in two different elections and thus must be excluded from the analysis.

It would be desirable if electoral data contained information on the number of new and lost voters between two elections in each unit. However, because such data are not usually available, the most reasonable strategy is to check the absolute relative change in the total number of voters within each unit: if this is below a given threshold and the two elections are close in time, one can simply adjust the data for the second election so that the total number of voters equals that of the first election and remove those units with an absolute relative change above the threshold. This is equivalent to assume that the new voters behave according to a vector of transition probabilities which is a weighted average if the remaining voters, an assumption which can be expected to do little damage as long as the proportion of new voters is small. When estimation is attempted for a large area or a whole country and local units are aggregated within larger administrative boundaries, it will be difficult to check the quality of the data. In addition, it is unlikely that the underlying assumptions are satisfied. As a consequence, the estimated transitions obtained in this way do not provide a consistent estimate of the average transitions for the whole country, even if one had access to accurate data, which is more difficult.

When, like in the Italian system, there is a large number of competing parties and some of them obtain a very small number of votes, two difficulties arise: (i) the normal approximation may not hold when applied to sparse table, (ii) due to the large number of parameters to be estimated, there will be a loss of efficiency and the transitions from small parties will be estimated with large standard errors. In such cases some aggregation of very small parties will be necessary; i any case one should fit a model with as many parties as possible and, like in the application above, aggregate and rescale at the end.

References

Anastasi, A:, Gangemi, G., Pavsic, R. Tomaselli, V. (1989) *Guerra dei flussi o bolle di sapone ?*, Bonannino, Acireale.

Bernardo, J. M. (2001) Interpretation of electoral results: a bayesian analysis. Internal report, Departamento de Estadistica i I.O., Universitat de Valencia.

Bracalente B., Ferracuti L., Forcina A. (2006) L'analisi dei flussi elettorali in Umbria: le elezioni dal 2004 al 2006. AUR&R, 7, 145-178.

Brown, P. J. and Payne, C. D. (1986) Aggregate data, ecological regression and voting transitions. *Journal of the American Statistical Association*, 81, 453-460.

Forcina, A. and Marchetti, G. M. (1989) (Modelling transition probabilities in the analysis of aggregate data. In *Statistical Modelling*, Decarli, A., Francis, B. J., Gilchrist, R., Seber, G. U. H. Eds. Springer Verlag.
Goodman, L. A. (1953) Ecological regression and the behaviour of individuals. *American Sociological Review*, 18, 351-367.
King, G., Rosen, O. and Tanner, M. A. (1999) Beta-binomial hierarchical models for ecological inference. *Sociological Methods & Research*, 28, 61-90.
King, G., Rosen, O. and Tanner, M. A., (Eds) (2004) *Ecological inference*. Cambridge University Press, Cambridge.
Plackett, R. L. (1977) The marginal totals of a 2 x 2 table. *Biometrika*, 64, 37-42.
Robinson, W. S. (1950). Ecological Correlations and the Behavior of Individuals. *American Sociological Review* 15: 351-357.
Wakefield, J. (2004) Ecological inference for 2×2 tables. J, of the Royal Statist. Soc. A, 167, 1-42.

Appendix

Assume that n_{iu} = $(n_{iu1},\dots,n_{iuC_{iu}})' \sim$ Mult$(n_{iu}, 1/C_{iu})$, that $y_{iuc} \sim$ Mult(n_{iuc}, p_{iuc}) and finally that $p_{iuc} \sim$ Dirichelet(θ_i, π_i). Direct calculations show that

$$\mathrm{Var}\,(y_{iu} \mid n_{iu}) = n_{iu}\Omega\,(\pi_i)\left[1+\theta_i\left(1+\sum_c n^2_{iuc}/n_{iu}-1\right)\right];$$

in order to compute the marginal variance note that $\mathrm{e}\,(y_{iu} \mid n_{iu})$ is simply $n_{iu}\Omega\,(\pi_i)$, this implies that $\mathrm{Var}\,[\mathrm{e}(y_{iu} \mid n_{iu})]$ is a null matrix and $\mathrm{Var}(y_{iu})$ = $\mathrm{e}\,[\mathrm{Var}(y_{iu} \mid n_{iu})]$ requires only to compute $\mathrm{e}(n'_{iu}n_{iu}$; by a well known result on expectations of quadratic forms,

$$\mathrm{e}(n'_{iu}n_{iu} = n_{iu}[1+(n_{iu}-1)/C_{iu}];$$

by substitution

$$\mathrm{Var}(y_{iu}) = n_{iu}\Omega\,(\pi_i)[1+\theta_i(n_{iu}-1)/C_{iu}]$$