Analysis of regional aspects of voting behaviour – the case of polish presidential election

Mariusz Mazurkiewicz[1]

Wroclaw University of Technology

keywords:
ecological regression,
electorate flows,
transition of votes,
homogeneity of electorate,
decomposition.

Some individuals change their political preferences each elections, even changes appear during one multi round election. Transition is observed when a voter who voted for a given candidate in first round of elections, votes for another candidate in final round of elections. Exact information about the scale of transitions is usually unavailable. There are many opinion pools, even exit pools, but exact transition's data is not collected during election. Ecological regression techniques give opportunity to obtain quantitative description of electoral behaviour from aggregated data. Aggregated data set is published after election. Ecological regression approach give reliable results under homogeneity assumption. Homogeneity in this case is considered in the term of electoral behaviour. For the whole country homogeneity assumption is usually insufficient. Regional decomposition of estimation process, used for the description of a voter's behaviour, extends ecological regression to application for large regions or even for the whole country. In this analysis maximum likelihood approach and regional decomposition of voting results is used. The last presidential election 2010 is used as empirical example in the paper. Presidential election in Poland may have one or two rounds. The time between rounds of election in Poland is two weeks.

Introduction

Voting behaviour expressed by voting results may be generally described by voting data at the individual level. Anonymous voting eliminates availability of this type of statistical information. On the other side, aggregated data available from published statistics, contains the information about individual voting behaviour. Information about individuals is expressed as results of group decision process. There exists a common approach in social sciences aimed at using aggregated data to analyse individual behaviour of voters. Data aggregation bases usually on geographic units such as countries and constituencies.

Our main goal is to obtain the proportions of voters who voted for the same candidate or who changed their preferences in a candidate's choice between stages of elections.

[1] Contact address: Institute I-23, Wroclaw University of Technology, Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland, e-mail: mariusz.mazurkiewicz@pwr.wroc.pl

Ecological regression model (Goodman, 1959) is one of the statistical tools which gives solution of this problem. The method is popular and well known especially in the case of two-party system. There are also many improvements, modifications and extensions of ecological regression method (Füle, 1994), (King 1997) (Mazurkiewicz *et al*. 2006). In our approach, we consider presidential election. We would like to obtain proportions from aggregated election data and estimate individual-level transition coefficient from electoral data aggregated over various voting districts.

The 2 x 2 case for consecutive elections

The main idea of ecological regression will be illustrated on two parties and two separate elections approach. It is obvious that replacement candidates instead parties makes this approach able to use for generally personal election, even presidential.
The simplest case in ecological regression approach is the 2 x 2 case. This case occurs when there are only two competitive parties in two consecutive elections: party *1* and party *2*.
Let N_{ij} ($i,j = 1,2$) describe the number of those who vote for party i in the first election and party j in the second election. In table 1 the situation is presented in more convenient way where $N_{.j}$ and $N_{i.}$ are marginal values for the first and the second elections respectively: $N_{.j}$ – part of electorate, which votes for the party j ($j=1,2$) in the second elections (it doesn't matter which party was supported by this part of the electorate in the first elections), and $N_{i.}$ – part of electorate voting for the party i in the first election (it doesn't matter which party will be supported by this part of the electorate in the second elections).

Table 1. Distribution of the electorate between two parties in two consecutive elections.

N_{11}	N_{12}	$N_{1.}$
N_{21}	N_{22}	$N_{2.}$
$N_{.1}$	$N_{.2}$	

For two parties and two consecutive elections all marginal values are known as a result of elections and all cellular values could be obtained as a solution of the system of equations. However, the 2x2 case is a simplification; it is not too difficult to show that even such a model is useful in practical analysis. If party *1* is a fraction of electorate taking part in the elections and party *2* means the part of the electorate not participating in these elections the above model is proper to use for evaluation of electorate's flows from absence to participation and vice versa. The participation or non-participation is frequently under investigation of policy makers. The quantitative description of electorate's flows gives complementary information about dynamic of the electorate in the sense of politically active part of the society.

In real life, the size of electorate is changing slightly for two consecutive elections. To simplify, the assumption that $N_{.1} + N_{.2} \cong N_{1.} + N_{2.} \cong N$ is sufficient. Under these assumptions the process of searching cellular values from table 1 is quite easy. Let x_i denote proportion of votes obtained by party i in the first election. Thus, $x_i = \frac{N_{i.}}{N}$, and $y_j = \frac{N_{.j}}{N}$ where y_j denotes proportion of votes obtained by party j in the second election. Let t_1 denote the proportion of voters who voted for the same party in the first election and in the second election. Let t_2 denote the proportion of voters who switched to another party between two consecutive elections. The results of the above transformation are presented in table 2.

Table 2. Transition of the electorate between two parties in two consecutive elections.

			Total
	t_1x_1	$(1-t_1)x_1$	x_1
	t_2x_2	$(1-t_2)x_2$	x_2
Total	y_1	y_2	*1*

According to the above notation, the proportion of voters who vote for party *1* in the second elections is a linear combination of x_1 (called ecological regression):

$$y_1 = t_1x_1 + t_2x_2 = t_1x_1 + t_2(1-x_1) = (t_1 - t_2)x_1 + t_2$$

The solution of such a classical model is well known (Duncan & Beverley, 1953); (King, 1997); (Groofman & Merril, 2002).

Generally, the relation between proportions of votes in the first elections and in the second elections is described by the system of the following regression functions:

$$y_1 = t_{11}x_1 + t_{12}x_2 + \varepsilon_1$$

$$y_2 = t_{21}x_1 + t_{22}x_2 + \varepsilon_2 \; ,$$

where $\varepsilon_1, \varepsilon_2$ are unobserved disturbances. Thus, normalized 2x2 case of transition of votes between elections is presented as in table 3.

Table 3. Coefficient of transition of votes between parties in two consecutive elections.

	t_{11}	t_{12}
	t_{21}	t_{22}
Total	1	1

The n x 2 case for consecutive elections

Voting results in two consecutive elections can be expressed as a cross-tabulation of election data. Let N denote the size of the whole electorate. The situation can be described in general like a description presented in table 4.

Table 4. Distribution of the electorate in multi round election (consecutive elections).

round I	round II				Total (round I)
	cand. 1	cand. 2	…	cand. q	
cand. 1	N_{11}	N_{12}	…	N_{1q}	$N_{1\bullet}$
cand. 2	N_{21}	N_{22}		N_{2q}	$N_{2\bullet}$
…	…	…	…	…	…
cand. p	N_{p1}	N_{p2}	…	N_{pq}	$N_{p\bullet}$
Total (round II)	$N_{\bullet 1}$	$N_{\bullet 2}$	…	$N_{\bullet q}$	N

In table 4 contingency tables notation is used. Marginal frequencies in the last row correspond to aggregated election results for the second round. Marginal frequencies in the last column correspond to aggregated elections results for the first round. Cellular frequencies N_{ij} denote the number of voters who voted for candidate *i* during the first round and for candidate *j* during the second round.

Results of both rounds are also described by the vector of votes given for parties taking part in first or second elections. The aim is to describe a transition-voting process in two consecutive elections with the probabilities of changes of political preferences. In this case, estimating procedure for transition probabilities is necessary when we analyse behaviour of the whole electorate. A classical ecological approach bases on the assumption that the results of the second election are a linear function of results of the first election:

$$TX + \varepsilon = Y \, , \qquad (1)$$

In above equation, T is a transition matrix, X is a vector of results of the first round, Y is a vector of results of the second round, and ε is a vector of unobserved disturbances.

$$T = \begin{bmatrix} t_{11} & t_{12} & \dots & t_{1k+1} \\ t_{21} & t_{22} & \dots & t_{2k+1} \\ \vdots & \vdots & \dots & \vdots \\ t_{k+11} & t_{k+12} & \dots & t_{k+1k+1} \end{bmatrix}; \quad X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{k+1} \end{bmatrix}; \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{k+1} \end{bmatrix},$$

where

x_j – share of votes for candidate j obtained during the first round;
y_j – share of votes for candidate j obtained during the second round;
k – total number of candidates (k=max(p,q));
x_{k+1} – abstention during the first round;
y_{k+1} – abstention during the second round;
t_{ij} – transfer coefficient of votes transferred from candidate j during the first round to votes for candidate i in the second round, i,j=1,2,..,k;
t_{k+1j} – transfer of votes for candidate j in the first round to abstention in the second round;
t_{ik+j} – transfer of votes from abstention in the first round to votes for candidate i in the second round.

Extended form of general equation

$$TX + \varepsilon = Y$$

is given by the system of equations:

$$\begin{cases} t_{11}x_1 + t_{12}x_2 + \dots + t_{1k+1}x_{k+1} + \varepsilon_1 = y_1 \\ t_{21}x_1 + t_{22}x_2 + \dots + t_{2k+1}x_{k+1} + \varepsilon_2 = y_2 \\ \vdots \\ t_{k+11}x_1 + t_{k+12}x_2 + \dots + t_{k+1k+1}x_{k+1} + \varepsilon_{k+1} = y_{k+1} \end{cases},$$

where coefficients t_{ij} fulfil conditions

$$\sum_{i=1}^{k+1} t_{ij} = 1$$

for all j=1,2,..,k+1.

In above system of equations transfer coefficients are unknown. Transfer coefficients have to be estimated from voting results for considered aggregation level. In this approach complex knowledge about voting results is assumed.

First of all, any estimation approach needs a statistical sample. In the case of ecological regression there is one way to obtain a statistical sample: by using aggregated voting results. Usually, the voting results data are available, even on basic electoral districts level. Assumptions that data from every electoral district are homogenous are obviously not held. There is a necessity to divide electoral districts into homogenous groups - in the sense of electoral behaviour. This approach will be called decomposition. The main idea of decomposition is to construct the decomposition with respect to reasonable assumption that small regions are more homogeneous than large regions. It is very convenient approach, because usually electoral data are divided into geographically selected voting districts. Homogeneity simply means that those transfer coefficients are the same or roughly the same for selected voting districts. This definition of homogeneity is difficult to evaluate in the sense of fitness, because transfer coefficients are unknown. Thus, the problem of formal definition of homogeneity of transfer electorate coefficients should be avoided.

Procedure of estimation transfer coefficients based on decomposition approach and assumption about electoral behaviour homogeneity in small voting district. For instance maximum likelihood methodology, which guarantees high statistical quality of results can be used.

Multi round election

Presidential election in Poland from 1991 generally contains one or two rounds. Election is finished after first round then one of candidates achieved 50% plus one vote. He or she is the winner of election. If there is no winner in first round, in second round there are only two candidates selected from each candidates by number of votes rank and simple majority is enough to win election. The more interesting situation we can observe then 2 rounds are necessary to elect the president and this type of situation is analysed. Time period between rounds is two weeks. It means that voters who voted for lost candidates in first round should transfer vote in second round to another candidate or decide to skip second round.

To build mathematical model as a version of equation 1 let assume that the total number of candidates in presidential election is equal to n. First round is the pre-selection type procedure and “winners” of first are candidates: candidate 1 has the biggest number of votes in first round, candidate 2 has the second biggest number of votes.

Mathematical model of vote’s transfers is described by system of equations. This system of equations contains three equations:

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n + a_{1n+1}x_{n+1} \\ y_2 = a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n + a_{2n+1}x_{n+1} \\ y_1 = a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \cdots + a_{3n}x_n + a_{3n+1}x_{n+1} \end{cases}$$

where

x_1 – share of votes obtained by candidate 1 in first round,

x_2 – share of votes obtained by candidate 2 in first round,

x_3 – share of votes obtained by candidate 3 in first round,

⋮

x_n – share of votes obtained by candidate n in first round,

x_{n+1} – abstention in first round,

y_1 – share of votes obtained by candidate 1 in second round,

y_2 – share of votes obtained by candidate 2 in second round,

y_3 – abstention in second round,

a_{ij} – transfer coefficients.

Results of first round fulfil condition

$$x_1 + x_2 + x_3 + \cdots + x_n + x_{n+1} = 1$$

and for every i=1,2,3,...,n

$$\frac{x_i}{1 - x_{n+1}} \leq \frac{1}{2}$$

Candidates 1 and 2 qualified to the second round. In terms of shares of votes $x_1 > x_i$ and $x_2 > x_i$ is hold for every i=1,2,3,...,n.

For the second round, condition

$$y_1 + y_2 + y_3 = 1$$

is obviously hold.

Transfer coefficients fulfils conditions:

$$\begin{cases} a_{11} + a_{21} + a_{31} = 1 \\ a_{12} + a_{22} + a_{32} = 1 \\ a_{13} + a_{23} + a_{33} = 1 \\ \vdots \\ a_{1n} + a_{2n} + a_{3n} = 1 \\ a_{1n+1} + a_{2n+1} + a_{3n+1} = 1 \end{cases}$$

and for all i=1,2,3 and j=1,2,3...,n+1 $a_{ij} \geq 0$.

Let assume that shares of votes for candidates 1 in first round is bigger than shares for candidate 2, $x_1 > x_2$.

Transfer coefficients have specific interpretation in double rounds election. For instance coefficient a_{11} is share of voters who supported the same candidate in both rounds, similar situation is observed for coefficient a_{22}. Both coefficients are recognized as a measures of electorate stability. Coefficients a_{12} and a_{21} contain information about the shares of voters who completely change preferences in two weeks period between rounds. It is probably influence of very intensive electoral campaign in last two weeks before final round. Let say it is measure of efficiency of last two weeks campaign. Coefficients a_{13}, a_{14}, ..., a_{1n} for candidate 1 and a_{23}, a_{24}, ..., a_{2n} for candidate 2 describe “pure” transfers from lost candidates to the winners of first round.

Empirical example

Decomposition approach was used to calculate common transition matrix for whole country. The simple administrative division was used as a base of regional decomposition. In this approach country is divided to 16 voivodeships and for each voivodeships separately transition matrix is estimated.

Table 5. Transfer coefficients for presidential election 2010 in Poland

i	a_{1i}	a_{2i}	a_{3i}
1	0,9942	0,0000	0,0058
2	0,0014	0,9966	0,0021
3	0,2541	0,5713	0,1747
4	0,6010	0,0441	0,3549
5	0,1699	0,7367	0,0933
6	0,4627	0,2293	0,3080
7	0,5033	0,1677	0,3290
8	0,6845	0,0337	0,2818
9	0,0925	0,8237	0,0838
10	0,3035	0,4433	0,2532
11	0,0200	0,0734	0,9066

.

The electorate stability in the term of two main candidates (participants of second round) is very high and similar. Parameters a_{11} and a_{22} from table 5 have almost the same values. Over 99% of electorate candidate 1 and 2 votes for them again in the second round. Over 90% of voters didn’t take part in both rounds. The electorate flow from abstention to participation is lower than 10% (coefficient $a_{3\,11}$, table 1). The rest of parameters describes natural vote’s flows from lost candidates to the second round participants, winners of first round.

Candidate number 7 had third result in first round. Transfer of voters, who supported him in first round for whole country is described by coefficients a_{17}, a_{27} and a_{37}. Parameter a_{17} is share of voters of candidate 7 in first round, who voted for candidate 1 in second round, analogously a_{27} is a share of voters who voted for candidate 2 and a_{37} is a share of voters decided to skip second round.

Figure 1. Flows of candidate 7 electorate.

Elements of common transition matrix are calculated as weighted averages of elements of regional transition matrices. Regional differences appear in comparison of results from voivodeships.

Table 6. Stability coefficients for sub-regions.

Sub-region	a_{11}	a_{22}	$a_{3\ 11}$
1	1,0000	1,0000	0,9344
2	0,9775	1,0000	0,9424
3	1,0000	1,0000	0,8683
4	1,0000	0,8772	0,9502
5	0,9857	1,0000	0,8370
6	1,0000	1,0000	0,8872
7	0,9797	1,0000	0,8416
8	1,0000	0,8772	0,9502
9	1,0000	1,0000	0,8988
10	1,0000	1,0000	0,9028
11	1,0000	0,9997	0,9833
12	1,0000	1,0000	0,9591
13	1,0000	1,0000	0,8296
14	1,0000	1,0000	0,8874
15	1,0000	1,0000	0,9322
16	1,0000	0,9550	0,9985

Regional coefficients for candidate 7 (third place in first round).

Table 7. Flows of electorate of candidate 7.

Sub-region	a_{17}	a_{27}	a_{37}
1	0,3265	0,3248	0,3487
2	0,5714	0,0897	0,3389
3	0,6574	0,0000	0,3426
4	0,6424	0,2282	0,1294
5	0,3854	0,0000	0,6146
6	0,4455	0,1735	0,3810
7	0,4540	0,0000	0,5460
8	0,6424	0,2282	0,1294
9	0,8419	0,0000	0,1581
10	1,0000	0,0000	0,0000
11	0,6196	0,3804	0,0000
12	0,5542	0,3346	0,1112
13	0,5224	0,0000	0,4776
14	0,2055	0,2440	0,5504
15	0,2986	0,2470	0,4544
16	0,6169	0,3831	0,0000

In many opinions and analysis the problem of turnout is very significant in term of final result. Common opinion in the circle of politician's campaign strategy advisors base on the statement that the key to win election is in light poll for some candidate or heavy poll for another. From this point of view the nonvoting group of electorate in first round and their behaviour in second round is very significant. In 2010 election about 9,5% decide to take part only in the second round. 2 % of them voted for candidate 1, the winner, 7,34% voted for candidate 2.

Table 8. Turnout and its flows – regional differences

Sub-region	$a_{1\ 11}$	$a_{2\ 11}$
1	0,0378	0,0278
2	0,0018	0,0558
3	0,0093	0,1224
4	0,0076	0,0422
5	0,0265	0,1365
6	0,0206	0,0923
7	0,0246	0,1338
8	0,0076	0,0422
9	0,0000	0,1012
10	0,0093	0,0879
11	0,0167	0,0000
12	0,0117	0,0292
13	0,0137	0,1567
14	0,0702	0,0425
15	0,0335	0,0343
16	0,0000	0,0015

Literature

1. Duncan, D., Beverley, D., 1953. *An Alternative to Ecological Correlation*. American Sociological Review vol. 18, 665-666.
2. Fűle, E., 1994. *Ecological Inference on Voting Data*. Department of Statistics, University of Lund.
3. Goodman, L., 1959. *Ecological Regression and the Behaviour of Individuals.* American Sociological Review vol. 18, No. 6 (December), 663-664.
4. Groofman, B., Merrill S., 2002. *Ecological Regression and Ecological Inference.* Prepared for presentation at the Ecological Inference Conference, 2002. Harvard University, Cambridge. http://course.wilkes.edu/merrill/.
5. King, G., 1997. *A Solution to the Ecological Inference Problem*. Princeton University Press.
6. Mazurkiewicz, M., Mercik J. Turnovec F., 2004, *Regional decomposition in Ecological Regression*, Proceedings of the 15th International Conference on Systems Science, Wroclaw 2004.
7. Mazurkiewicz M., Mercik J. Turnovec F., 2006 *Transfer of electorate's preferences during 2001 and 2005 parliamentary elections in Poland* (in Polish), [in:] „Badania operacyjne i systemowe 2006. Metody i techniki". (J. Kacprzyk and R. Budziński – eds). Warsaw, "Exit", 2006, 175-183.
8. Mazurkiewicz M., Mercik J. Turnovec F., 2010, *Transition of votes for consecutive elections – mathematical modelling.* Paper presented on annual meeting of European Public Choice Society, Efez, April 2010.

Appendix

Results of presidential election 2010 in Poland as a shares of population.

Candidate	Name	Round I	Round II
1	KOMOROWSKI Bronisław Maria	22,6571%	28,9742%
2	KACZYŃSKI Jarosław Aleksander	19,8885%	25,6832%
3	JUREK Marek	0,5755%	
4	KORWIN-MIKKE Janusz Ryszard	1,3530%	-
5	LEPPER Andrzej Zbigniew	0,6966%	-
6	MORAWIECKI Kornel Andrzej	0,0701%	-
7	NAPIERALSKI Grzegorz Bernard	7,4640%	-
8	OLECHOWSKI Andrzej Marian	0,7868%	-
9	PAWLAK Waldemar	0,9550%	-
10	ZIĘTEK Bogusław Zbigniew	0,0959%	-
11	Abstention	45,4575%	45,3426%
-	Turnout	54,5425%	54,6574%

Source: www.pkw.gov.pl