Ecological Inference of Regional Voting at Each District

Youngjae Jin
Yonsei University, Political Science
134 Shinchon-dong, Seodaemun-gu, Seoul,
Korea. y2jin@yonsei.ac.kr

Kihong Eom
University of Kentucky, Political Science
1645 Patterson Office Tower, Lexington, KY, 40506,
USA. keom0@pop.uky.edu

Subtitle: The 16th National Assembly Election of Korea

The aim of this paper is to disaggregate the vote share of candidate into the proportion of regional vote conditional on the percentage of hometown at *each* district and to examine whether regional voting pattern exists outside of political party' turf. Cross-level inference has encountered difficulties, because of "ecological fallacy" or "aggregation bias," which creates biased and inefficient estimates (Palmquist 1993). Several solutions for ecological fallacy have been introduced over years (Voss 2000; King, Rosen, and Tanner. 1999; Burden, and Kimball 1998; King 1997; Benoit, and King 1996; Achen, and Shively 1995; Palmquist 1993).[1] So far, utilizing two approaches (method of bounds and statistical approach), King's ecological inference outperforms over other methods (King 1997; King, Rosen, and Tanner 1999), because the estimate is most efficient and robust under the violation of assumptions.[2] The estimation of King's method takes two steps; the first step is to find the bounds of parameters, β^p and β^o. The second step is to estimate β^p and β^o at each district borrowing strength from other districts, viz., the truncated bivariate posterior distribution of βi^p and βi^o. The estimation of λ^h and λ^o is sequentially performed in the 2 by 3 table drawn from 2 by 2 table.[3]

1.1 The first step:

	Vote	No vote	Marginal
Hometown	βi^h	$1-\beta i^h$	X_i
Other Regions	βi^o	$1-\beta i^o$	$1-X_i$
Marginal	T_i	$1-T_i$	

1.2. The second step

	Party	Other party		No vote	Marginals
Hometown	λi^h	$1-\lambda i^h$	βi^h	$1-\beta i^h$	xi
OtherRegions	λi^o	$1-\lambda i^o$	βi^o	$1-\beta i^o$	1-xi
	P		Ti		

[1] . Simple introduction of previous models may begin with Goodman model. The equation would be $Ti = \beta^h Xi + \beta^o(1-Xi) + e$ (note that β is constant). Homogeneous model has the same equation system with Goodman model, but the β^h and β^o are calculated from the observed homogeneous unit. The third would be informed assumption model. With the same equation of Goodman model, it assumes we know some of parameters, viz., β^h or β^o. Therefore, we just estimate one of parameters assuming β^h (or β^o) is equal to zero. Final model would be neighborhood model. This model has the opposite view of Goodman model. It assumes $\beta^h = \beta^o$, but varies over districts. The equation would be $Ti = \beta iXi + \beta i(1-Xi) + e$, where $\beta i = a + bXi$ (see Steve 2000). Above models are motivated to solve the identification problem in the ecological regression (see Palmquist 1993). The existing models have two limitations to perform, however, let alone inefficient and less robust estimate; first, they do not enable to estimate the parameter of interest in this paper (λ_i^h). They just estimate β_i^h. The second limitation is that they cannot estimate more than 2 by 2 table.

[2] . Three assumptions are single mode of parameter of interest, the absence of spatial autocorrelation, and no correlation of marginals and parameter of interest. In the case of the violation of the last assumption, the estimate will be robust if the bounds are sufficiently narrow for many of lower level units (King, Rosen, and Tanner 1999, 67-68, but see Cho forthcoming).

[3] . Variables are weighted by the number of voting age at each district.

Where β and γ are the estimated parameter. Whereas capital Xi (the proportion of vote share) is fixed, small xi is not fixed because xi is equal to $\beta_i X_i / T_i$.[4]

The equation we attempt to estimate is to regress the vote share of each candidate at each district on the percentage of hometown. In each step, the information from deterministic bounds and ' borrowing strength' is utilized. The equation would be: $Pi = \lambda_i^h * x_i + \lambda_i^o * (1-x_i) + e$, where $xi = \beta_i X_i / T_i$ and $\beta i \sim$ Binomial ($\beta i \mid T$). P is the vote share of candidate, X is the percentage of hometown, and i means district. The program used is EzI, which is operated on Gauss procedure (version 2.4. 2/21/2001 release). The number of districts analyzed is 106 out of 227 districts in the 16th National Assembly election of Korea on April 2000: Seoul (29), Busan (9), Daeku (3), Inchon (8), Kwangju (3), Ulsan (2), Kyung-ki (19), Kangwon (2), Chungbok (6), Chunbok(3), Chunnam (5), Kyungbok (7), and Kyungnam (10).

All in all, disaggregating the regional voting conditional on the percentage of hometown, we locate two findings. The first one is that Seoul, Kyung-ki, Inchon, and Kangwon do not appear as regional voting bloc (λ^h) which refers to supporting the candidates who came from his/her hometown (Shin, Jin, Gross, and Eom 2001). One of persistently alleged characteristics in Korean politics has been regional voting. The relatively small proportion of voting is shown as regional voting, however [e.g., in Kyung-ki region, .3805 for the Grand National Party, .4309 for the National Congress for New Politics and .1231 for the United Liberal Democrats]. The second one is that there is a tendency of ' regional assimilation' to attenuate the myth of strong regional voting bloc such as Youngnam and Honam in South Korea. If they were born in Honam region, but are living in Youngnam region, the regional voting for the hometown party of Honam is .1506. The case of voters from Youngnam living in Honam region is .041.[5]

REFERENCE

Achen, Christopher H. and W. Phillips Shively. 1995. *Cross-Level Inference*. Chicago: University of Chicago press.

Benoit, Kenneth and Gary King. 1996. "A Preview of *EI* and *EzI*: Program for Ecological Inference." *Social Science Computer Review* 14:433-438.

Burden, Barry C. and David C. Kimball. 1998. "A New Approach to the Study of Ticket Splitting." *American Political Science Review* 92: 533-544.

Cho, Wendy K. Tam. Forthcoming. "Iff the Assumption Fits…A Comment on the King Ecological Inference Solution." *Political Analysis*.

King, Gary, Ori Rosen, and Martin A. Tanner. 1999. "Binomial-Beta Hierarchical Models for Ecological Inference." *Sociological Methods & Research* 28:61-90.

King, Gary. 1997. *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data.* Princeton, NJ: Princeton University Press.

Palmquist, Bradley Lowell. 1993. *Ecological Inference, Aggregate Data Analysis of U. S. Elections, and the Socialist Party of America*. Ph. D. Dissertation.

Shin, Myungsoon, Youngjae Jin, Donald A. Gross, and Kihong Eom. 2001. "Money Matters in Party-Centered Politics: Campaign Spending in Korean Congressional Elections." Working paper.

Voss, D. Stephen. 2000. *Familiarity Doesn't Breed Contempt: The Political Geography of Racial Polarization*. Ch. 6. Ph. D. Dissertation.

[4] . Note that due to sequential division of cell the more columns has, the more the uncertainty of the estimate increases (King 1997, ch 8.)

[5] . One may argue this is because of high absence rate of those people, who were born in one region but are living in other region. Therefore, this phenomenon could be good news or bad news.