Using a Bayesian Hierarchical Probit Model to determine the Effects of Social Class on Voting Behavior

Armijo, Roberto.

Introduction

This paper is an attempt to bridge the gap between the rational choice theory approach to voting behavior, specifically the spatial theory of voting [see Downs 1959], with the at first apparently incompatible sociological approach; through the use of hierarchical Bayesian model, I will attempt to examine the effects of belonging to a social group affects the assumptions of spatial voting.

Among the approaches to study voting behavior are rational choice theory and the sociological approach [see Lijphart 1979 and 1980]; both of these approaches have instrumental underpinnings. While the sociological approach focuses on the impact of social structure or how belonging to a social group affects voting behavior as individuals vote for the party that best reflects the interests of their social class, the rational choice approach assumes voting behavior to be more individualistic; assuming the individual's voting behavior is the product of an individual cost-benefit analysis.

Model and Data

Although these approaches may seem incompatible at first glance, there are a number of ways to close the gap between them; one of these possibilities is given by the fact that the rational choice approach allows for the specification of social class identity in the individual's utility function. The approach of this paper is different in the sense that it does not alter the basic utility function but rather though a Bayesian Hierarchical Probit model, it allows the variation of parameters according to social class.

The data used for this paper corresponds to the Chilean *Centro de Estudios Públicos* (CEP) survey of October of 2009[1], the last survey taken before the December election, which resulted in a run-off between right-wing coalition candidate Sebastian Piñera, and the candidate for the incumbent left-wing coalition, Eduardo Frei.

$$U_{ji} = \beta * z_i + \varepsilon_{ji}$$

The first step is defining the standard spatial model model. Being U_{ji} a representation of voter's i utility if candidate j is elected, which is a function

[1] Centro de Estudios Públicos. "*Estudio de Nacional de Opinión Pública, Oct. 2009*". *Website www.cepchile.cl*

dependent on ideological position; where $z_i$represents the ideal point for the i[th] voter as determined by the Cahoon-Hinich method of factor analysis, this is done through the analysis of thermometer feeling scores on the CEP survey.

$$T_{ji} = -(|\pi_j - z_i|)^2 + \varepsilon_{ji}$$

Being T_{ji} the i^{th} respondent's score on the j^{th} politician; π_j the coordinates of the j[th] candidate; and z_i the ideal point for the i[th] voter and ε_{ji} being random error. By taking the survey data, the method estimates z_i and π_j and assuming the thermometer scores are linear in quadratic Euclidian distance.

In the two-candidate run-off election, voters should vote for candidate j (1,2) if the utility they deride from that candidate's election is larger than the utility derided from the opponent winning.

$$\Pr[\text{vote} = 1] = \Pr[U_{1i} > U_{2i}]$$

By taking this model and letting the parameters of the probit regression vary by social class, I attempt to determine how this variable affects voting behavior. In order to achieve this, I use the Bayesian hierarchical probit model represented below to compute the posterior density for the parameters via a Gibbs sampler.

```
model{

for(i in 1:553){
y[i] ~ dbern(pi[i])
probit(pi[i])<-alpha[class[i]]+beta[class[i]]*x[i]


}

for(j in 1:3){
alpha[j] ~ dnorm(MeanA, TauA)
beta[j] ~ dnorm(MeanB, TauB)
}
MeanA ~ dnorm(0, .01)
MeanB ~ dnorm(0, .01)
SigmaA ~ dunif(0, 100)
SigmaB ~ dunif(0, 100)
Sigma ~ dunif(0, 100)
TauA <- pow(SigmaA, -2)
TauB <- pow(SigmaB, -2)
Tau <- pow(Sigma, -2)
}
```

This approach allows avoiding the problems of estimating the parameters independently for each social class such as a small amount of data for each class or little meaningful variation within each class and those of complete pooling. *"Hierarchical models deal with the possibility of variation across groups by positing an model for the parameters. The 'hierarchy' arises because the model for the parameters sits 'above' the model for the data."* (Jackman 2009, p.302).

Results

For the sake of comparison, the results of a GLM probit model are represented in in figure 1. These results allows us to observe that as individuals' ideological position is further to the right, the probability that they would vote for Sebastian Piñera in the run-off election increases.

Figure.1: GLM Probit model for probability of vote choice

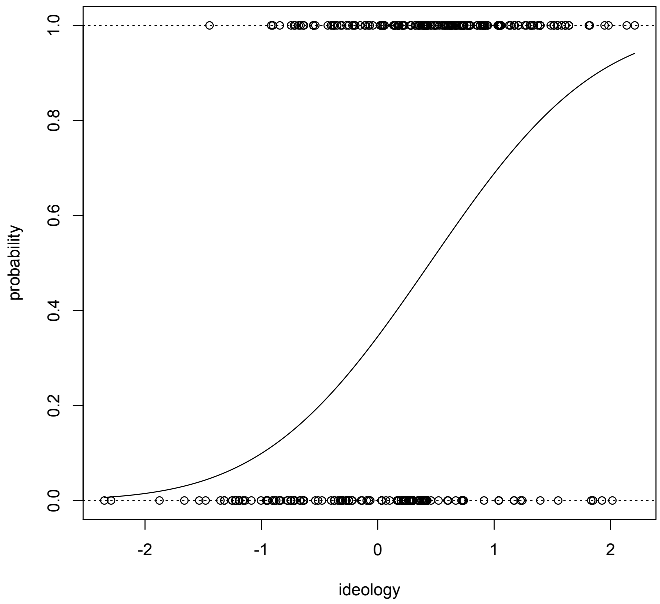
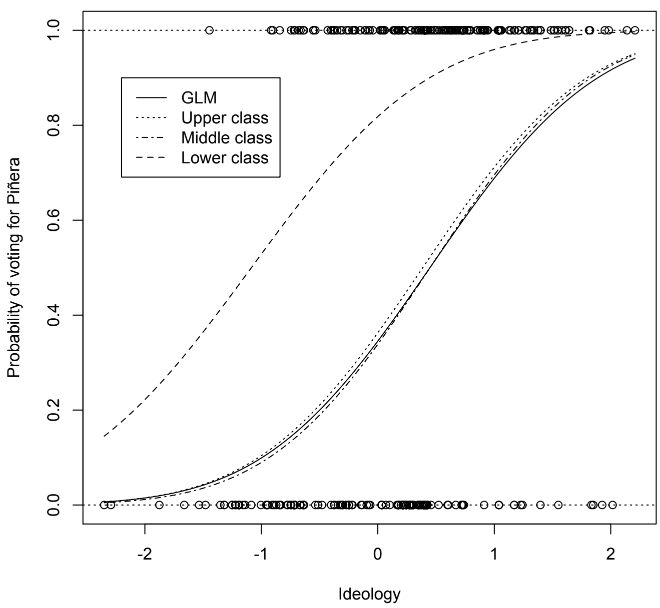


Figure 2, shows the results of the Hierarchical Probit model as well as those of the previously mentioned GLM model. Interestingly there is not much variation from the previous model for upper and middle class individuals but there is a much higher probability for lower class individuals to vote for the right wing candidate across the ideological spectrum. For further analysis of these results see the diagnostics appendix.

Figure.2: Hierarchical probit model of vote choice



Conclusion

On this paper I've examined the effects social class had on voting behavior on the 2009 Chilean elections, the results show that social class alters vote choice In these elections. Surprisingly, the results show that lower class individuals are had a higher probability of voting for the right wing candidate than upper or middle class individuals.

These results are preliminary and call for further analysis as lower class individuals would, according to the sociological approach, be more likely to vote for left-wing parties as they would better represent their interests. There is also the question on how political information is distributed across the population.

This paper shows the advantage of using the Bayesian approach, as allowing the parameters to vary by social class offers a mid-point alternative to either estimating the model group by group or complete pooling, both approaches that have problems of their own as there is an inverse relationship between bias and variance on which both of these approaches push to the extremes of each.

REFERENCES:

Centro de Estudios Públicos. "*Estudio de Nacional de Opinión Pública, Oct. 2009*". *Website* www.cepchile.cl

Downs, Anthony. "An *Economic Theory of Political Action in a Democracy*" The Journal of Political Economy 65:135-150. 1957.

Jackman, Simon. Bayesian Analysis for the Social Sciences. Wiley. 2009.

Lijphart, A. "Religion vs. Linguistic vs. Class Voting," American Political Science Review 65:686. 1979.

Lijphart, A. "Language, Religion, Class, and Party Choice: Belgium, Canada, Switzerland and South Africa Compared," In R. Rose (ed.), Electoral Participation: A Comparative Analysis. Beverly Hills: Sage. 1980.

APPENDIX: DIAGNOSTICS

Gibbs sampling is a common form of MCMC simulation in Bayesian analysis; as stated by Jackman (2009, p.252), MCM algorithms initialized at arbitrary starting values will eventually generate samples from the posterior density. The question is whether the algorithm has run long enough for the samples being drawn correspond to a valid characterization of the desired posterior density. A good practice to avoid using "bad samples" is throwing out the first iterations of the sampler, in this case I allow for a burn-in period of 5000 iterations after which, I allow the sampler to run 100,000 further samples. The 5000 burn in appears to be longer than necessary as the samples seem to converge a few hundred iterations in.

A preliminary method of assessing the convergence of the sampler on the target posterior density is that of graphically inspecting the output of the MCMC algorithm. Figures 3 and 4 show the traceplot and density for the sampled parameters of the probit regression; the traceplot seems to suggest convergence and the density functions approach normal bell curve shapes.

Figure 5 represents the autocorrelation functions of the parameters for the probit model, the fast decay in the autocorrelation functions seem to suggest that the sampled values should not have problems with stationarity.

Figure 3: Traceplot and density for intercepts

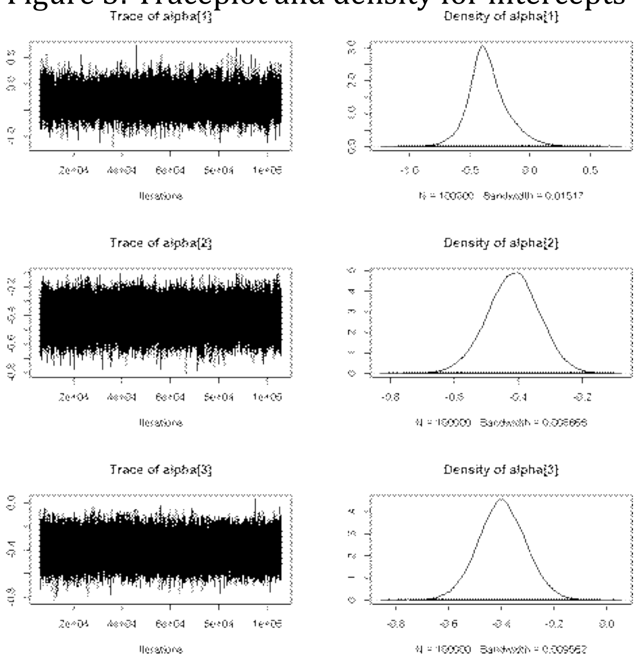


Figure 4: Traceplot and density for slopes

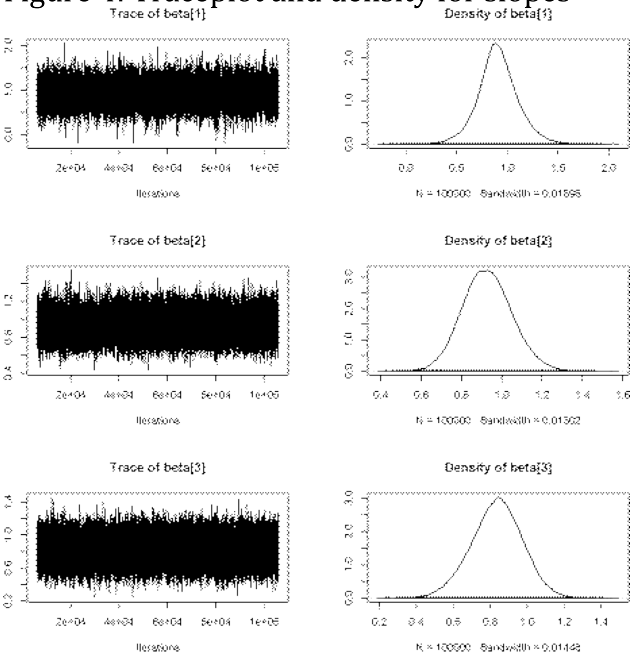


Figure 5: Autocorrelation factors for intercepts and slopes

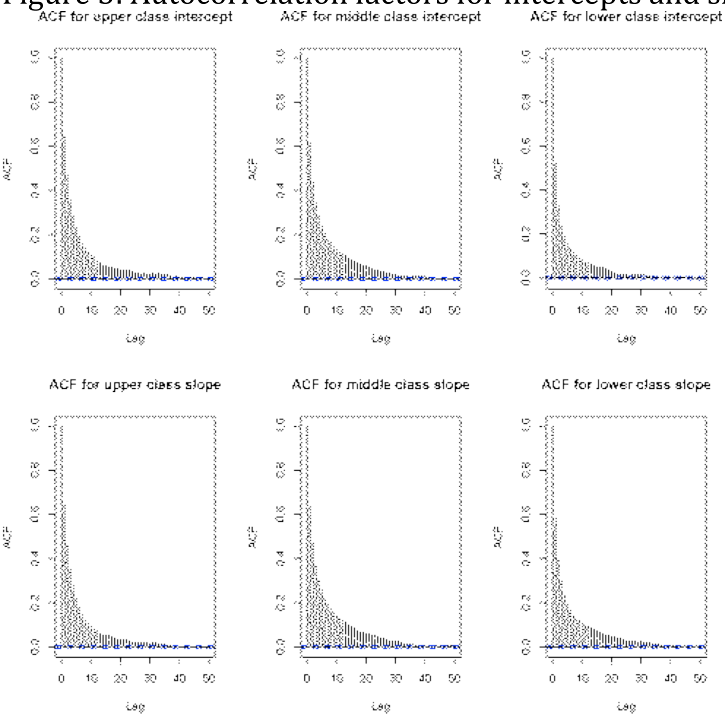


Table 1 shows less a subjective convergence diagnostic in the Geweke test, which compares the means of MCMC output of the parameters at two different stages of the MCMC run; in this case the first 10% compared to the last 50%. The t-values of the Geweke test for the parameters suggest there isn't a significant difference in the MCMM stages, which can be interpreted as convergence. Additionally, the test on effective size for each parameter shows over 10 thousand usable samples for each, meaning the samplers have run long enough.

Table 1: Stationarity test and effective size of parameters

Coefficient	**Geweke test**	**Effective size**
Lower class intercept	1.4136	11573.29
Middle class intercept	0.2688	11049.87
Upper class intercept	1.1794	15506.25
Lower class slope	-0.6286	13090.03
Middle class slope	0.1232	10378.16
Upper class slope	1.2373	12375.43