Application of the Variable Choice Logit Model to the British General Elections of 2010

Elena Labzina and Norman Schofield

October 6, 2014

Abstract

.

Keywords. varying choice logistic model, Bayesian statistics, British General elections 2010

In most democratic countries the same sets of parties run for the elections in all national constituencies. This is the main reason why the set-up of a single domain of the votes' alternatives may be justified. This is exactly the case of the canonical version of the formal stochastic vote model (Schofield and Sened. 2006; Schofield, 2007), where the multinomial logistic approach (MNL) is applied to the spatial multi-dimensional ideological framework. However, not in all countries every party competes in all constituencies. By definition, therefore, the necessary assumption of the irrelevant alternatives (IIA) (e.g Austen-Smith and Banks, 2000) may not be claimed. Hence, MNL may not be used any more. Among developed countries, the most straight-forward instances requiring an adjustment for MNL are Canada and UK. In Canada, *Bloc Quebecois* runs for the elections only in Quebec (for more details, see McAlister et al, 2013), while in the UK, the influential *Scottish National Party* competes only in Scotland,

and *Plaid Cymru* - in Wales[1]. The case of UK is considered in this chapter.A study of Scotland is particularly relevant because of the referendum on Scottish independence on September, 2014. The method deployed here is also relevant in many countries in Europe where there are regional parties, including Spain, Belgium and Italy.

Fortunately, a theoretical solution for the problem of the violation of the IIA has been already proposed. In his paper, Yamomoto (2011) describes a modification of the logistic MNL model to serve this purpose, called the varying choice logit (VCL) model. Also, he provides an actual example of its application to the elections in Japan. The idea of this adjustment is to estimate the categorical probabilities based only on those alternatives from which an individual could actually choose. For completeness, the probabilities of unfeasible options are explicitly set to zero. This approach naturally splits the voting sample based on their available choice sets of parties.

In contrast to Yamamoto, who programs the maximum likelihood method explicitly, this paper takes the Bayesian tools provided by the R package *rjags*, applying the Metropolis algorithm (e.g Gelman et al, 2014) to the data on the British General Elections of 2010. Indeed, as Gallego et al (2014) note, a natural solution to the problem with the random effects is the hierarchical Bayesian model. Despite being heavy "computationally", it provides a useful means to control for the unobserved randomness. Especially, given that as the technologies develop the time costs decrease, however, still being higher that those of the "direct" estimation.

Methodologically, this work can be considered a more complex continuation of McAlister et al (2013). In that paper, the authors look at Canada, where in Quebec an additional influential party (Bloc Quebecois), runs for the elections. Meanwhile, in Britain both Wales and Scotland have their own specific parties

[1] Northern Ireland has completely its own parties and is excluded from this analysis

that collected a significant share of votes in those regions during the General Elections of 2010. The Scottish National Party (SNP) gained 19.9% in Scotland, being the second after the Labour Party, which gave it about 2.6% of the total national vote. Compared to that, the share of Playd Cymru (PC) in Wales does not look impressive: only 0.6 at the national level - some other small parties managed to get even more - however, in Wales PC got 11.14 %, which clearly suggests the violation of IIA.

Substantively, our paper is the next step after Schofield et al(2011), where the British General Elections of 2005 and 2010 are analyzed separately for England, Wales and Scotland. Their results for 2010 show an unexpected insignificance of the valence of the Labour Party, which is easily explained, given the results of this work. The valence on the national level becomes significant (still very small), when the structure of the bundle of the parties in Scotland, where the Labour Party had a tremendous support of 42% in 2010, is properly accounted for in the analysis. This satisfies the expected assumption that the Scottish electorate influenced the electoral strategy of the Labour party during the General Elections of 2010.

First, the paper makes an overview of the extended formal model (Yamomoto (2011)) with the focus on the particular case of the UK. Second, the model is applied to the data from the British Election Study 2010. Third, the results of the substantive priors, the newly obtained estimates and the counterfactuals for the valences of the parties are compared briefly. Lastly, the assessment of the convergence of the party positions is performed according to the specification presented in detail in McAlister et al (2013).

This work introduces a pure spatial electoral model with regional additions to the valences. Some of the findings to be presented are:

1. the VCL model reveals that the counterfactual model underestimates the

spatial effect;

2. the valence of the Labour party is very close to zero;

3. the proper inclusion of Scotland decreases significantly the estimated valence of the Conservative Party.

Furthermore, the paper proposes an approach that can be easily extended in the further research on UK or applied to other countries, where IIA is not met.

1 Formal model in the application to the UK

This is a modification of the canonical Schofield's analysis (e.g Schofield and Sened 2006) for the specific case with the varying individual choice sets of the parties. As it was already mentioned before, because of the violation of IIA, the application of the usual multinomial logistic regression is impossible.

In UK, three major parties - the Conservative, Labour and Liberal Democrats - gained 88.2% of the votes in the General Elections of 2010. The rest of the votes were split among minor parties (at the national scale), which could have been neglected in the following analysis if all voters had the same alternatives in their electoral choice bundles. Our major interest is, first, the investigation of the party valences, and, second, the convergence of the positions, especially those of the major parties. The challenge emerges from those two parties that each ran for the elections only in one specific region: the Scottish National Party in Scotland and Playd Cymru in Wales. In simple words, because of this we cannot assume that an individual in Wales did not not vote for SNP, simply because he did not want, since he did not have such an option at all. Important to note, that the implicit assumption made is the similarity of all voters across the regions, except for the sets of the voting alternatives.

In the further analysis, we assume that the full set of parties consists of the five parties labeled with the numbers from 1 to 5: Labour(1), Conservative(2), Liberal Democrats(3), SNP(4) and Plaid Cymru (5). Then, denoting the region of individual i[2] as r(i) and the utility i obtains from voting for party j as $u_{i,j}$. if x_i is the political position of the individual and z_j is the position of j party, three possible sets of the utilities exist in the analysis:

$$r(i) = 1 \implies \bar{u}_{i\,r(i)} = \{1, 2, 3\} \implies u_i(x_i, z|r(i) = 1) = \{u_{i\,1}, u_{i\,2}, u_{i\,3}\}$$

$$r(i) = 2 \implies \bar{u}_{i\,r(i)} = \{1, 2, 3, 4\} \implies u_i(x_i, z|r(i)) = 2) = \{u_{i\,1}, u_{i\,2}, u_{i\,3}, u_{i\,4}\}$$

$$r(i) = 3 \implies \bar{u}_{i\,r(i)} = \{1, 2, 3, 5\} \implies u_i(x_i, z|r(i) = 3) = \{u_{i\,1}, u_{i\,2}, u_{i\,3}, u_{i\,5}\}$$

In words, it means that given the region of an individual his or her personal utility is modeled only for the parties for which this person could vote. Particularly, In England (1), only three major parties (1-3) competed at the elections. Meanwhile, in Scotland(2) and Wales(3) SNP(4) and PC(5) ran respectively as well.

The Bayesian approach enables to include extra random intercepts to capture the regional variation of the valences for the three major parties ($\mu_{i\,j},\ i = 1, 2,\ j = 2, 3$). Here, England is assumed as the base region and the Liberal Democrats - as the base party. We do not include regional valences for SNP and PC in the model, since they ran only in one region.

Therefore, the general form of the most straight-forward version of the individual utility function (e.g Schofield. and Sened (2006)):

$$u(x_i, z_j) = \lambda_j - \beta||x_i - z_j|| + \mu_{jr(i)} + \epsilon_i \tag{1}$$

[2] 1 for England, 2 for Scotland, 3 for Wales

Here, λ_j is the valence of party j, which is the intercept specific for party j, and β is the common (for all parties) spatial coefficient, representing the personal sensitivity to the ideological deviation of the party from the individual "bliss" point. Hence, x_i is the individual ideological position, while z_j is the ideological position of party j. $\mu_{jr(i)}$ is the regional addition to the valence specific for region r(i)'s support for party j, which is the region of the individual.

Assuming that only for some combinations of j and r(i) the regional-party valences are significantly different from zero, e.g. $E(\mu_{jr(i)}) \neq 0$, the aggregated *mixed valence* of party j, with the control for the number of the individuals from each region, may be defined as:

$$\lambda'_j = \lambda_j + \frac{1}{n} \sum_{r(i):E(\mu_{jr}) \neq 0} n_r \mu_{jr} \tag{2}$$

This is an innovation in relation to the canonical definition of valences. This definition takes into account the distribution of the potential electorate of a party across the regional units. For instance, given the equal size of the regional subsamples, the second component on the right-hand becomes an average of the regional valences. Meanwhile, in the case with a base party k, its regional effect is set to 0.

Based on the definition above, now we can define the *mixed valences* for the five parties of interest as:

$$\lambda'_1 = \lambda_1 + \frac{1}{n}(\mu_{12} n_2 + \mu_{13} n_3)$$

$$\lambda'_2 = \lambda_2 + \frac{1}{n}(\mu_{22} n_2 + \mu_{23} n_3)$$

$$\lambda'_3 = 0$$

$$\lambda'_4 = \lambda_4$$

$$\lambda_4' = \lambda_5$$

Finally, a conservative[3] way to estimate the standard error of the mixed valence is proposed:

$$sd(\lambda_j') = \sqrt{Var(\lambda_j) + \frac{1}{n^2} \sum_{r(i):E(\mu_{jr})\neq 0} n_r^2 Var(\mu_{jr})} \tag{3}$$

1.1 Predicted vote shares

Another important point is the approach to predict the sample probabilities, if one wants to compare the sample and the model in terms of the predicted votes. To remind, in this canonical setting the predicted probability (and the share of votes) is $\hat{p_j} = \sum \hat{p_{ij}}/n$, where $\hat{p_{ij}}$ is the individual prediction for the probability of the vote for party j.

Meanwhile, in the VCL model, in which the estimated individual probability is $p_{i\,j} = \frac{e^{u_{i\,j}}}{\sum\limits_{k\in m(i)} e^{u_{i\,k}}}$, where $m(i)$ is the set of parties from which i chooses, the prediction of the votes for j in the region of i, $r(i)$, is

$$\hat{p_{j\,r}} = \frac{\sum\limits_{i\in r} p_{i\,j}}{n_r}$$

Hence, the total probabilities are:

$$\hat{p_j} = \frac{1}{n} \sum_{r\in m(i)} n_r \, \hat{p_{j\,r}}$$

Here, n_r is the number of the observations for region r, and n is the total size of the sample. It is easy to be that n_j's sum up to one.

In a way, these probabilities, considering the full set of 5 parties, are tentative. No individual in the sample chooses from all of them. However, these

[3] Assuming no correlation between the terms, which most probably leads to the overestimation of the magnitude, and requires refinement in the further work on this topic.

probabilities can be understood as a general characteristic of the each party, the expected vote share, that incorporates two probabilities: the weight or "the probability" of the region and, given the region, the probability to vote the party. To generalize:

$$p_j = \int p(j|r)p(r)dr$$

The conservative estimates (assuming the zero correlation of the probabilities across the regions) of the standard errors for the probabilities on the national level:

$$sd_{p_j} = \sqrt{Var(p_j)} = \sqrt{\frac{1}{n^2} \sum_{r \in m(i)} n_r^2 Var(p_{j\,r})}$$

1.2 Convergence at the critical point

As McAlister et al (2012) have shown in their paper[4], the combination of the party positions, $z = (z_1, ..., z_k)$, is the critical point in the model if:

$$\frac{\partial V_j(z)}{\partial dz_j} = \frac{2\beta}{N_j} \sum_{t=1}^{w} \sum_{i=1}^{n} (x_{it} - z_{jt})\rho_{ij}(1 - \rho_{ij}) = 0$$

Here $V_j(z)$ is the votes maximized by party j and w is the number of the ideological dimensions.

Then, a critical point is a Local Nash Equilibrium (LNE), or, in simple words, a set of positions from which each party has no rational grounds to deviate, given the proximate radius, the following conditions must be satisfied:

I. All eigenvalues for the matrices of the second derivatives (the Hessian) with respect to z must be negative.

[4] To see more mathematical details on the derivation of the first and second other conditions for V_z, see McAlister et (2012), pp. 6-15

The Hessian for the varying choice sets and the 2-dimensional policy space[5]:

$$H_{j1}(z_j^*) = \begin{pmatrix} \frac{2\beta}{n'} \sum_{i=1}^{n'} p_{i\,j}(1-p_{i\,j})(2\beta(x_{i1}-z_{j1})^2(1-2p_{ij})-1) \\ \frac{4\beta^2}{n'} \sum_{i=1}^{n'} (x_{i2}-z_{j2})(x_{i1}-z_{j1})p_{i\,j}(1-p_{i\,j})(1-2p_{ij}) \end{pmatrix} \tag{4}$$

$$H_{j2}(z_j^*) = \begin{pmatrix} \frac{4\beta^2}{n'} \sum_{i=1}^{n'} (x_{i2}-z_{j2})(x_{i1}-z_{j1})p_{i\,j}(1-p_{i\,j})(1-2p_{ij}) \\ \frac{2\beta}{n'} \sum_{i=1}^{n'} p_{i\,j}(1-p_{i\,j})(2\beta(x_{i2}-z_{j2})^2(1-2p_{ij})-1) \end{pmatrix}$$

Important to note, n' is not the total number of the observations in the sample, but the number of the individuals having party j in their choice bundle, since a party is expected to maximize only across those who can vote for it. Hence, for our case $n'_{LAB} = n'_{CON} = n'_{LAB} = n_1 + n_2 + n_3 = n$, $n'_{SNP} = n_2$ and $n'_{PC} = n_3$.

2. The convergence coefficient of the electoral system, which is the largest convergence coefficient among the parties, must be less than 1 with respect to each policy dimension.

Therefore, in the 2-dimensional case, the convergence coefficient for each party:

$$c_j(z) = c_j^1(z)+c_j^2(z) = \frac{2\beta}{n'} \sum_{i=1}^{n'} p_{i\,j}(1-2\,p_{i\,j})(x_{i1}-z_{j1})^2 + \frac{2\beta}{n'} \sum_{i=1}^{n'} p_{i\,j}(1-2\,p_{i\,j})(x_{i2}-z_{j2})^2 \tag{5}$$

$c_j^1(z)$ and $c_j^2(z)$ are the components related to each dimension of the policy.

Then, for party j:

- If $c_j(z) > 2$ then there is no convergence.
- If $c_j(z) < 2$, $c_j^1(z) < 1$ and $c_j^2(z) < 1$ then the system converges.

[5]Based on McAlister et al (2013), pp 13-14

The convergence coefficient for the whole electoral system:

$$c(z) = max(\{c_j(z)\}) \tag{6}$$

If conditions I and II are satisfied, then each party has no incentives to deviate, given that other parties do not deviate. Formally, this means that z is a Local Nash Equilibrium for the electoral system

On the next, computational stage of the analysis, various modifications of the equation (1) were tried. Based on the parsimonious grounds, the final model assumes non-zero $\mu_{jr(i)}$'s for $(j, r(i)) = \{(1,2),(2,2),(1,3),(2,3)\}$ only (while j=3, Liberal Democrats, is the reference party).

2 The British General Elections in 2010

2.1 Variables of the analysis

This paper uses the data from British Election Study 2009-2010. The individuals from Northern Ireland were excluded from the survey. Anywhere, further in the text Great Britain refers to England, Wales and England.

The variables of the analysis are:

1. Dependent variable: *the party voted* may take the following values:

    - Labour
    - Conservatives
    - Liberal Democrats
    - Scottish National Party
    - Plaid Cymru

2. Independent variables: *Survey questions used to construct the individual scores of the ideological dimensions of nationalism (anti-EU) and economy (anti-taxes)* [6].

3. Control variable: *the region of the respondent:*

   - England
   - Scotland
   - Wales

The intent was to keep as many observations as possible. Hence, only those observations that had explicit 'Don't know' in the independent questions were dropped. Control and dependent variables had no missing values. These missing independent variables were assumed *missing at random*(MAR). To fill in the missing values the R package MICE (Multivariate Imputation by Chained Equations) was used [7].

2.2 Sample and electoral statistics

Table I and Table II present the comparison of the electoral results for the General Elections of 2010 in UK. Seemingly, among the parties of our interest Liberal Democrats and SNP are overrepresented in the sample, while the Conservatives and, especially, PC are underrepresented. This is the feature of the raw sample and, for example, Schofield et al (2011), who used the same survey, have very similar summary statistics.

[6] For more details on the questions see the Appendix of Schofield at al (2011). This paper employs the exact same set of the questions.

[7] This can be done with the sequential application of the *mice()* and *impute()* functions

Table 1. 2010 Election in Great Britain					
	Population			Sample[1]	
Party	Vote %	Seats	Seat %	Observations	Observations %
Conservative	36.1	306	47.1	3,097	35.43
Labour	29.0	258	39.7	2,350	26.89
Liberal Democrats	23.0	57	8.8	2,384	27.28
Scottish National Party	1.7	6	0.9	210	2.40
Plaid Cymru	0.6	3	0.9	43	0.49
Others	9.6	20	2.6	656	7.51
Total	100	650	100	8,740	100

[1] Based on 2010 British Election Survey campaign Internet panel data

Table 2. 2010 Election in Great Britain by Region: Voting									
	All			Scotland			Wales		
Party[1]	el[3](%)	obs[2]	obs (%)	el (%)	obs	obs (%)	el(%)	obs	obs (%)
Con	36.1	3,097	38.31	16.7	134	16.75	26.1	110	28.50
Lab	29.0	2,350	29.07	42.0	283	35.38	36.2	137	35.49
LibDem	23.0	2,384	29.49	18.9	173	21.63	20.1	96	24.87
SNP	1.7	210	2.60	19.9	210	26.25	-	-	-
PC	0.6	43	0.56	-	-	-	11.3	43	11.14
Total	90.4	8,084	100	97.5	777	100	93.7	370	100

[1] Only major parties and region-specific parties: Con: Conservative Party; Lab: Labor Party; LibDem: Liberal Democrat Party; SNP: Scottish National Party; PC: Plaid Cymru

[2]Sample based on BES 2010 containing only observations of those voted for the 5 parties

[3] Elections

2.3 Component factor analysis: Ideological positions

In the original dataset the ideology of the respondents is represented with 8 related questions. Meanwhile, the core independent variables in our spatial model are two ideological coordinates, the axis representing the attitudes towards *nationalism* and *economy*. Hence, we need "to shrink" the number of dimensions from 8 to 2, and component factor analysis (CFA) is the method to apply in this situation. As Gill writes, its "*basic idea to linear transform a dataset into smaller dimension dataset with the property that each of the transformed variables is uncorrelated*" (Gill, undefined).

This procedure is done with the R function *factanal* that performs maximum likelihood factor analysis. To minimize the interaction of the dimensions, the basis is set to be orthogonal, hence the *varimax* rotation is used in the computation (e.g Abdi, 2003).

Table 3 presents the results of our CFA analysis.

Table 3. 2010 Factor Analysis		
	Nationalism	Economy
1. EU membership	0.894	
2. EU cooperation	0.845	0.174
3. Nuclear plan	0.281	0.395
4. Tax-spend	-0.325	-0.388
5. Tax exemption		0.373
6. Mansion tax	0.118	0.632
7. Tax relief		0.294
8. Ecotax	0.266	0.392
n	8084	
% variance	0.223	0.140
Cumulative % Variance	0.223	0.363

The correlation matrix for Great Britain in 2010 is:

$$\nabla_0 = \begin{bmatrix} & nat & econ \\ nat & 0.867 & 0.066 \\ econ & 0.066 & 0.592 \end{bmatrix}.$$

Important to note that the ideological scores are weakly correlated: 0.066. Interestingly, the ideological preferences along the dimension of the nationalism are more spread out.

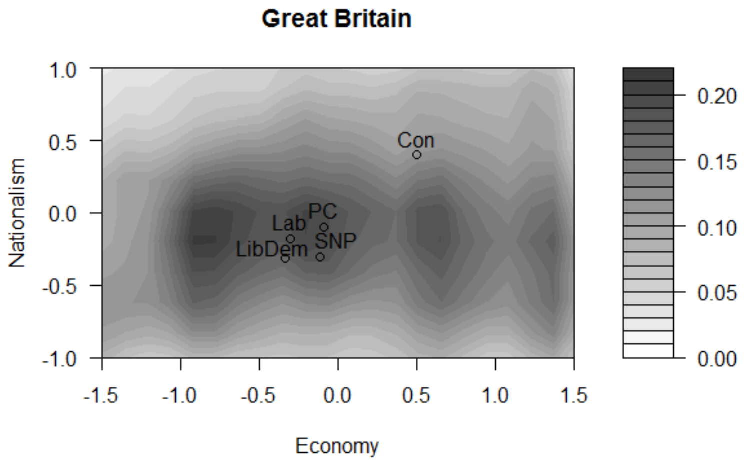


Figure 1. Density plot of the ideological dimensions: Great Britain 2010

The pattern of the spatial distributions shows that in contrast to the dimension of the nationalism, according to which the individuals are symmetrically distributed around the origin, the economic dimension provides 3 clear "tops", one of which is located around the origin, while two others lie on each side of the origin in about 0.7 points (Figure 1). If we look at the same plot for the case of England (Figure 2), we observe a very similar picture. Meanwhile, Scotland (Figure 3) provides only one extreme point, corresponding to one of those from the aggregated plot. In Wales (Figure 4), two extreme point present.

Interestingly, in this case no dense area exists proximate to the origin. Meanwhile, the "left" concentration "top" is practically the same as that unique one in Scotland.

The ideological individual averages grouped by the party voted:

$$z^* = \begin{array}{c} \\ nat \\ econ \end{array} \begin{array}{c} \begin{array}{ccccc} Lib & Con & LibDem & SNP & PC \end{array} \\ \begin{pmatrix} -0.31 & 0.50 & -0.34 & -0.11 & -0.09 \\ -0.18 & 0.40 & -0.31 & -0.31 & -0.10 \end{pmatrix} \end{array}$$

These results provide the evidence that *the Conservatives* is the only party positive in terms of their ideological coordinates. Meaning, they are more anti-EU and more anti-tax than the average across the population. This observation may be, probably, at least, a partial explanation of their success at the General Elections of 2010. Furthermore, the supporters of *the Conservatives* are the most extreme in terms of the magnitude. Among the pro-EU and pro-tax supporters, *the Liberal Democrats* are the most extreme. The supporters of *Plaid Cymru* are the least extreme.

2.4 Spatial model

The canonical model[8] modified to include the assumption of the varying party bundles can be specified as:

$$u(x_i, z_j) = \lambda_j - \beta||x_i - z_j|| + \mu_{jr(i)} + \epsilon_i \tag{7}$$

Here, for individual i the utility is assumed to be defined only for those parties for which he or she can actually vote. The major innovation is the regional valence, $\mu_{jr(i)}$. Since significantly more observations come from England than from Scotland and Wales, 6898 of 8084, England is set to be the "base" region

[8] See, for example, Schofield and Sened, 2006

in the model. Consequently, in terms of the coefficients, the assumption is $\mu_{jr(i)} \neq 0$ only if $(j, r(i)) \in \{(1,2)(2,2)(1,3)(2,3)\}$, while $\mu_{jr(i)} = 0$ for the rest, where j $\in$ {1,2,3,4,5} $\equiv$ {*Labour*, *Conservatives*, *LibDem*, Scottish National Party, Plaid Cymru}) and r(i) $\in$ (1, 2, 3) $\equiv$ {*England*, *Scotland*, *Wales*}.

2.4.1 Non-informative or "substantive" priors?

The Bayesian approach requires specification of the prior distribution for the coefficients. For the model introduced above (7) this means setting priors to 12 random variables: one for each coefficient and three more to specify their dispersion (see Table 4). Given a large number of iterations, the role of the prior distributions is not very important, since the posterior distribution with the increase of the number of iterations converges to the true distribution.

First, for simplicity, each coefficient is assumed to be distributed normally with the dispersion having with the inverse-gamma distribution with the parameters 0.1 and 0.1. The only question left: what is the best way to specify the means for the normal distribution of the coefficients? Is there any difference if the posterior distribution, in any case, must converge eventually to the true distribution?

As it is going to be shown soon, in terms of the final estimates, small differences in the "starting" means do not affect the results. However, it is still interesting to compare the results given that some prior information is used or not.

What are the informative, "substative", priors for the means? Why can they be of interest? Let's imagine the situation in which researchers cannot apply this relatively complicated VCL approach. Meanwhile, they may not want to violate IIA explicitly. Then the best predictor for the valence of a party is the estimate from MNL, that was run for the largest region in which this party runs. This is very similar to the approach proposed by Yamamoto (2011) for choosing the starting points for ML to estimate VCL.

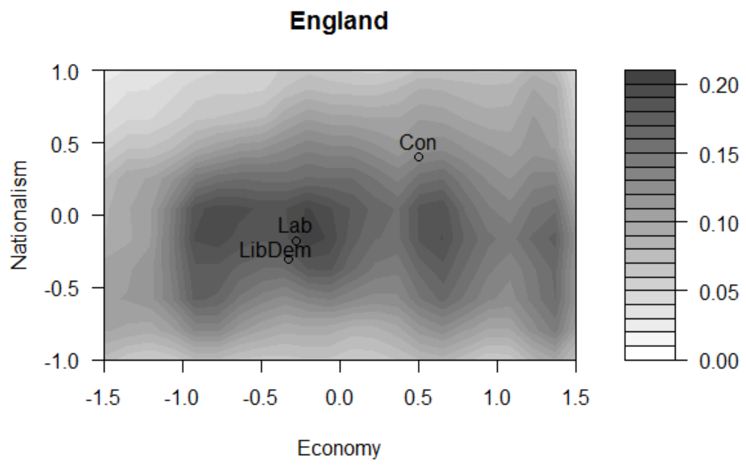


Figure 2. Density plot of the ideological dimensions: England

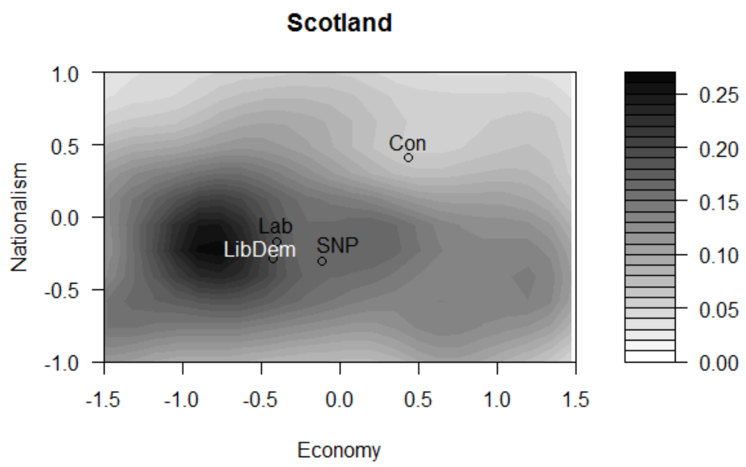


Figure 3. Density plot of the ideological dimensions: Scotland

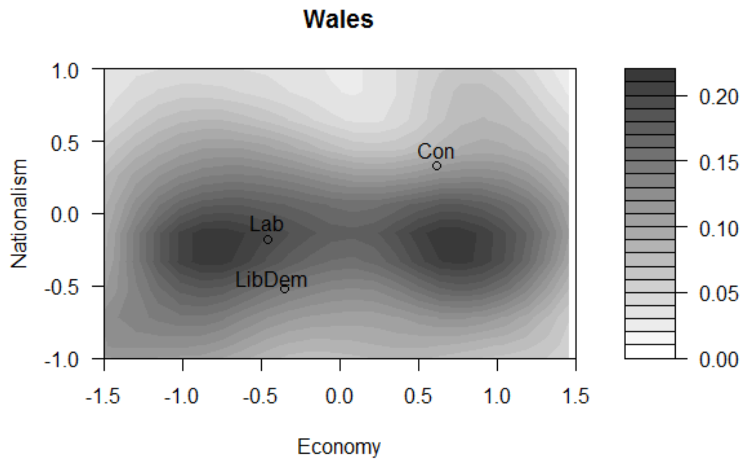


Figure 4. Density plot of the ideological dimensions: Wales

In case of this paper, we get the informative prior estimates for the Labour and the Conservatives from MNL run for England; for SNP for Scotland, and for PC for Wales. The ordered means for the prior distributions of the valences are:

$$\lambda^{prior} = (\lambda^{prior}_{PC}, \lambda^{prior}_{Lab}, \lambda^{prior}_{LibDem}, \lambda^{prior}_{SNP}, \lambda^{prior}_{Con}) = (-0.827, -0.099, 0, 0.233, 0.276) \tag{8}$$

Interestingly, it turned out that based on the prior estimates Plaid Cymru is the least attractive party, and the Conservatives, expectedly, is the most attractive.

Table 4 shows the specification for two sets of priors. As it can be seen, the distinction between them is the means for λ_1, λ_2, λ_4, λ_5.

Table 4. Prior distributions of the coefficients		
	Non-informative	Informative
β	$N(0, 1/\tau_1)$	$N(0, 1/\tau_1)$
λ_1	$N(0, 1/\tau_2)$	$N(-0.099, 1/\tau_2)$
λ_2	$N(0, 1/\tau_2)$	$N(0.276, 1/\tau_2)$
λ_4	$N(0, 1/\tau_2)$	$N(0.233, 1/\tau_2)$
λ_5	$N(0, 1/\tau_2)$	$N(-0.827, 1/\tau_2)$
λ_{21} λ_{21} λ_{31} λ_{32}	$N(0, 1/\tau_3)$	$N(0, 1/\tau_3)$

* $\tau_i \sim Gamma(0.1, 0.1)$

2.4.2 Bayesian estimation

Based on the above priors, two models are set. In both models, after running 25000 iterations of 3 chains with the R package *rjags*, the convergence is confident according to the Gelman-Rubin diagnostic: all partial potential scale reduction factors and the multivariate potential scale reduction factor equal to 1. In relation to the Heidelberger-Welch diagnostic both tests have passed in all chains and for all coefficients [9] (see the plots of the Gibbs sampling in the Appendix). The results of the estimated models are provided in Table 4. The estimates are expectedly very close across the models.

[9] with a minor exception in one chain for μ_{Lab3}, which is expected since the confidence interval contains 0

Table 5. Pure spatial model with regional effects				
	Model 1		Model 2	
	estimate	credible interval (95 %)	estimate	credible interval (95 %)
β	0.873	[0.833, 0.913]	0.873	[0.833, 0.914]
λ_{Lab}	-0.102	[-0.163,-0.039]	-0.101	[-0.163,-0.039]
λ_{Con}	0.259	[0.193, 0.325]	0.260	[0.194, 0.326]
λ_{SNP}	0.227	[0.024,0.432]	0.227	[0.025,0.432]
λ_{PC}	-0.762	[-1.127,-0.409]	-0.762	[-1.125,-0.408]
μ_{Lab2}	0.589	[0.391,0.792]	0.590	[0.391, 0.790]
μ_{Con2}	-0.466	[-0.729,-0.208]	-0.466	[-0.728, -0.207]
μ_{Lab3}	0.458	[0.193, 0.726]	0.459	[0.198, 0.722]
μ_{Con3}	-0.056	[-0.363,0.247]	-0.056	[-0.363, 0.250]
DIC	Mean/penalized deviance 15175/15184		Mean/penalized deviance 15175/15184	
N	8084			

Table 6 shows the results for the mixed party valences relative to the base Liberal Democratic party.

Table 6. Regional mixed valences				
	England	Scotland	Wales	Mixed
λ_{PC}	-	-	-0.762	-0.762
λ_{Lab}	-0.102	0.487	0.091	0.002
λ_{Con}	0.259	-0.470	0.622	0.213
λ_{SNP}	-	0.227	-	0.227

2.5 Counterfactuals and comparison

What would happen if we pretend that IIA holds? This is, probably, the most interesting question to answer, since it addresses the gain from the development of the VCL model, which is the aim of this paper.

In this section, the counterfactual results, the substantive priors from the previous section, and the VCL estimates are compared.

In the counterfactual model, the estimates for SNP and PC are expectedly highly negative: since most of the voters from our sample could vote for them, the assumption of the opposite must lead to the underestimation of the valences of SNP and PC. Hence, *a priori*, we know that those estimates should not make much sense. Meanwhile, the counterfactual estimates for the major parties potentially (as one can suppose) might be still making sense.

Table 7 presents the estimates from the counterfactual model, meanwhile, Table 8 provides three sets of the results (substantive priors, counterfactuals, and VCL).

Table 7: Counterfactual MNL	
base=LibDem	
Variable	Est
	(|t-stat|)
β	0.761***
	0.017
λ_{Lab}	-0.024
	0.029
λ_{Con}	0.281***
	0.031
λ_{SNP}	-3.903***
	0.084
λ_{PC}	-5.489**
	0.160
n	8084
LL	-8396.6
McFadden R^2	0.141

In Table 8, the most catching eye result is the estimate for the Labour Party which is slightly negative, while in the corresponding VCL model it is confidently positive. The source of that is a higher popularity in Scotland and Wales relative to that in England. This holds despite the presence of such close alternatives as Scottish National Party and Plaid Cymru, the effect of which is captured by the VCL model. Furthermore, the relative popularity of the Conservatives is 25% lower in the VCL model (0.28 and 0.21). This is a consequence of the structure of the voting preferences in England, where the Conservatives have the most of their support. The last feature to mention is the larger spatial coefficient in the VCL. Hence, the counterfactual model underestimates the votes' sensitivity to the policies (the spatial coefficient is 0.76 in the counterfactual model, while in the VCL it is 0.87).

Table 8: Comparison of the results

	(1)	(2)	(3)
β	-	0.87	0.76
λ_{lab}	-0.01	0.002	-0.02
λ_{con}	0.28	0.21	0.28
λ_{snp}	0.23	0.23	-3.90
λ_{pc}	-0.83	-0.76	-5.49

Models:

(1) - "Substantive" Priors

(2) - *Mixed* valences from the pure spatial VCL model

(3) - Counterfactuals

2.6 Assessment of the convergence at the origin

The last aim of our investigation is to check whether the voting system is stable if the parties take mean positions relative to their constituencies. Practically, this is performed via plugging in the numbers below (9) into the formulas from Section 1.2 using the VCL results from Section 2.4.2.

$$z^0 = (z^0_{lab}, z^0_{con}, z^0_{lib}, z^0_{snp}, z^0_{pc}) = \begin{pmatrix} 0 & 0 & 0 & -0.19 & -0.09 \\ 0 & 0 & 0 & -0.13 & -0.11 \end{pmatrix} \tag{9}$$

The conclusions of this subsection are especially interesting, being looked at as an examination of the two-dimensional variation of the median voter theorem.

The Hessian matrices for the parties are:

$$H_{Lab|z^*} = \begin{pmatrix} -0.185 & -0.008 \\ -0.008 & -0.222 \end{pmatrix} \quad H_{Con|z^*} = \begin{pmatrix} -0.266 & -0.031 \\ -0.031 & -0.275 \end{pmatrix} \quad H_{LibDem|z^*} = \begin{pmatrix} -0.203 & -0.034 \\ -0.034 & -0.238 \end{pmatrix}$$

$$H_{SNP|z^*} = \begin{pmatrix} -0.098 & 0.014 \\ 0.014 & -0.211 \end{pmatrix} \quad H_{PC|z^*} = \begin{pmatrix} 0.024 & 0.002 \\ 0.002 & -0.060 \end{pmatrix}$$

The estimates above show that for the major parties the relative costs of the departure from the origin have a trade-off between the dimensions: the cross-derivatives are negative. Interestingly, in the case of SNP, the cross-derivative is positive, hence a tiny potential synergy effect between the dimensions exists for this party. The Hessian for PC provides the evidence that it could do better by moving along the nationalism dimension positively.

$$eigen(H|z^*) = \begin{array}{cc} & \begin{matrix} Lib & Con & LibDem & SNP & PC \end{matrix} \\ \begin{matrix} Nat \\ Econ \end{matrix} & \begin{pmatrix} -0.183 & -0.240 & -0.182 & -0.096 & 0.024 \\ -0.224 & -0.302 & -0.258 & -0.213 & -0.060 \end{pmatrix} \end{array}$$

$$\{c(z^*)_j\} = \begin{pmatrix} c(z_1^*) \\ c(z_2^*) \end{pmatrix} = \begin{array}{cc} & \begin{matrix} Lib & Con & LibDem & SNP & PC \end{matrix} \\ \begin{matrix} Nat \\ Econ \end{matrix} & \begin{pmatrix} 0.115 & -0.201 & 0.080 & 0.178 & 0.126 \\ 0.084 & -0.239 & 0.057 & 0.089 & 0.072 \end{pmatrix} \end{array} =$$

$$\begin{array}{c} \begin{matrix} Lib & Con & LibDem & SNP & PC \end{matrix} \\ \begin{pmatrix} 0.2 & -0.44 & 0.137 & 0.267 & 0.197 \end{pmatrix} \end{array}$$

$$c(z) = max(\{c(z^*)_j\}) = 0.267 < 1$$

We observe that the only party for which the median position is the saddle point is Playd Cymru. Meanwhile, the rest of the conditions for the convergence of the electoral system are confidently satisfied. Dependent on our perception of PC, this situation may be evaluated in two different perspectives.

First, formally, this electoral system does not converge, if we consider all parties to be equally important. The second approach is to remember that the inclusion of Plaid Cymru was rather technical, mostly to control formally for the violation of IIA in Wales. However, clearly it is not a major player, and many parties not included in the analysis got significantly more voters (e.g Green Party).

What is the most important, in terms of the three major parties, the electoral system confidently converges. This does not change with the inclusion of the only other candidate that can be considered as a major party - Scottish National Party.

An interesting observation is the negative convergence coefficient for the Conservatives. This means that they benefit extremely from their position, and even a small deviation from the mean position might cause a significant decrease of the voting support.

3 Conclusion

This paper presents an example of the Bayesian application of the varying choice logistical model to the electoral data from the British General Elections of 2010. The British electoral system in 2010 was shown to converge at the origin considering the Labour, Conservative, Liberal Democratic, and Scottish Nation parties. However, it diverges with the addition of Plaid Cymru.

4 References

Abdi, H. (2003). Factor rotations in factor analyses. *Encyclopedia for Research Methods for the Social Sciences.* Sage: Thousand Oaks, CA, 792-795.

Austen-Smith, David, and Jeffrey S. Banks (2000). *Positive political theory I: collective preference. Vol. 1.* University of Michigan Press.

British Election Study 2010, *University of Essex*

McAlister, K., Jeon, J. S., and Schofield, N. (2013). Modeling elections with varying party bundles: Applications to the 2004 Canadian election. In *Advances in Political Economy,* Schofield, Caballero, G. and Kselman, D. (eds.pages 289–315. Berlin,Springer.

Gallego, M., Schofield, N.,and Jeon, J., and McAlister, K.(2014) The variable Choice set Logit Model applied to the 2004 Canadian election. *Public Choice* 158: 427-463.

Gelman, Andrew, Carlin, John B., Stern, Hal S., Dunson, David B., Vehtari, Aki, and Donald B. Rubin (2014). *Bayesian Data Analysis.* CMC Press.

Gill, Jeff (undefined), What is principle component analysis anyway? *Notes*

Schofield, N. (2007). The mean voter theorem: Necessary and sufficient conditions for convergent equilibrium. *Review of Economic Studies* 74: 965–980.

Schofield, N., Gallego, M., and Jeon, J. (2011). Leaders, voters and activists in the elections in Great Britain 2005 and 2010. *Electoral Studies,* 30(3):484–496.

Schofield, N.(2005). A valence model of political competition in Britain: 1992-1997, *Electoral Studies,*24(3):347-370.

Schofield, N. and Sened, I. (2006). *Multiparty democracy: elections and legislative politics*. Cambridge, Cambridge University Press.

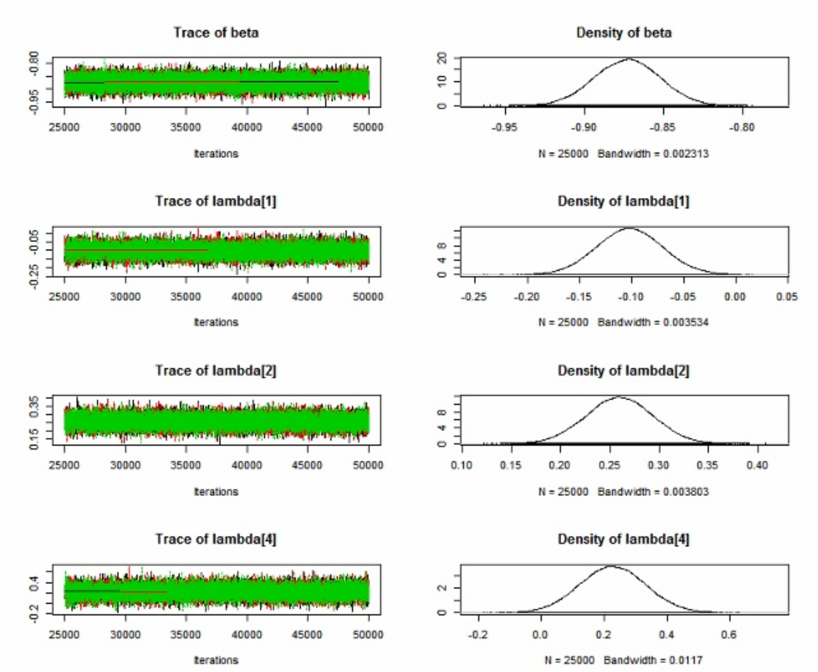
Yamamoto, T. (2011). A multinomial response model for varying choice sets, with application to partially contested multiparty elections.working paper.

5 Appendix

Table 9. Predicted and sample voting probabilities (Pure spatial model with regional effects)

	England			Scotland		
	sample	est	conf interval (95%)	sample	est	conf interval (95%)
p_{Lab}	0.280	0.280	[0.277, 0.283]	0.354	0.352	[0.346, 0.359]
p_{Con}	0.414	0.413	[0.407, 0.420]	0.167	0.167	[0.156, 0.182]
p_{LibDem}	0.307	0.307	[0.303, 0.310]	0.216	0.216	[0.211, 0.221]
p_{SNP}	-	-	-	0.262	0.263	[0.258, 0.267]
p_{PC}	-	-	-	-	-	-
	Wales			All sample		
p_{Lab}	0.355	0.352	[0.339, 0.365]	0.291	0.291	[0.291, 0.291]
p_{Con}	0.285	0.286	[0.261, 0.311]	0.383	0.383	[0.382, 0.384]
p_{LibDem}	0.250	0.251	[0.239, 0.262]	0.295	0.295	[0.294, 0.295]
p_{SNP}	-	-	-	0.026	0.026	[0.026, 0.026]
p_{PC}	0.111	0.112	[0.109, 0.114]	0.005	0.005	[0.005, 0.005]

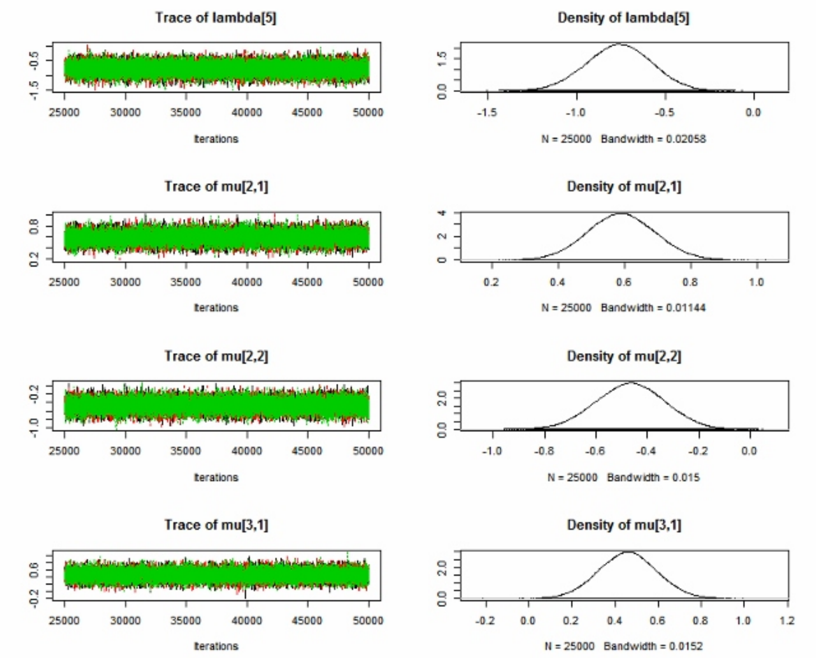
(1)



$_5.png$

Figure 1: Figure 5. Gibbs sampling for MCMC: Pure spatial models with regional effects 1

(1)



$_6.png$

Figure 2: Figure 6. Gibbs sampling for MCMC: Pure spatial models with regional effects 2

5.1 Code

Pure spatial model with the regional valences: Uninformative priors for the valences

```
basicVCL = function()
{
    for(i in 1:N)
   {
         for(k in 1:K)
            {
             v[i,k] <- lambda[k] + beta*((eu[i]-peu[k])^2 +
                                         (tax[i]-ptax[k])^2) + mu[region[i],k]
            expv[i,k] <- exp(v[i,k])*phi[region[i],k]
            pv[i,k] <- expv[i,k]/sum(expv[i,1:K])
            }
         vote[i] ~dcat(pv [i, 1:K])
   }
   lambda[1] ~dnorm(0, taul)
   lambda[2] ~dnorm(0, taul)
   lambda[3] <- 0
   lambda[4] ~dnorm(0, taul)
   lambda[5] ~dnorm(0, taul)
   beta ~dnorm(0,1/1000)
   for (p in 2:3)
   {
       for (y in 1:2)
      {
```

```
          mu[p,y] ~dnorm(0,taum);
         }
   }
   mu[p,3] <-0
   mu[p,4] <-0
   mu[p,5] <-0
   for (p in 1:5)
   {
       mu[1,p] <-0
   }
    for (y in 1:3)
   {
       for (k in 1:3)
      {
          phi[y,k] <-1
      }
   }
   phi[1,4] <- 0
   phi[1,5] <- 0
    phi[2,4] <- 1
   phi[2,5] <- 0
   phi[3,4] <- 0
   phi[3,5] <- 1
   taum ~dgamma(.1,.1);
   taul ~dgamma(.1,.1)
}
```