Estimating Constituency Opinion in Britain

Technical Report *

Chris Hanretty[†] Nick Vivyan[‡]

October 9, 2015

*This work was supported by Secondary Data Analysis Initiative grant from the Economic and Social Research Council [grant number ES/K003666/1]

[†] Reader in Politics, University of East Anglia. E-mail: c.hanretty@uea.ac.uk.

[‡] Lecturer in Quantitative Social Research, Durham University. E-mail: nick.vivyan@durham.ac.uk.

Contents

1 Introduction

For this research project we have estimated public opinion at the level of Westminster parliamentary constituencies for a number of topics. This report details the data and model used to estimate constituency opinion for each of these topics.

2 General estimation strategy

This section lays out the generic estimation strategy we use for our constituency opinion estimates. This strategy incorporates global smoothing, local smoothing, constituency level predictors, as well as individual level predictors and post-stratification. For a more in-depth discussion of the logic of each component of the statistical model, see our validation paper (Hanretty et al., 2014a).

For a binary political opinion variable y, our goal is to estimate the proportion of citizens with a score of $y = 1$ in constituency $j \in \{1, \ldots, 632\}$. We begin with a national survey of size N. For each respondent i in this survey, we have measures of constituency location, political opinion y_i, and $\{1, \ldots, K\}$ categorical demographic variables. Each of these respondent-level demographic variables takes on a value l_k from a set of L_k possible values.

In addition to the respondent-level variables, we also have information at the constituency level. First, we have a matrix of constituency-level variables X_j. Second, we have a square 632×632 spatial adjacency matrix whose elements $\omega_{jj\prime}$ are equal to one where constituency j and constituency $j\prime$ are geographic neighbours, and zero otherwise.

Step one: modelling political opinion

Based on this information, we model individual opinions y_i as follows:

$$\Pr[y_i = 1] = \text{logit}^{-1}(\alpha^0 + \alpha^1_{l_1[i]} + \alpha^2_{l_2[i]} + \ldots + \alpha^K_{l_K[i]} + \phi^{constituency}_{j[i]} + \upsilon^{constituency}_{j[i]}). \quad (1)$$

Here, α^0 is a grand intercept and $\alpha^k_{l_k[i]}$ is the effect of individual i being in category l_k of demographic variable k.[1] $\phi_j^{constituency}$ is a spatially autocorrelated constituency random effect whose distribution conditions on the value of $\phi^{constituency}$ in neighbouring constituencies. $\upsilon_j^{constituency}$ is a constituency random effect which is not spatially correlated, but which is modelled hierarchically as a linear function of constituency-level variables.

To model $\phi_j^{constituency}$ we follow Selb and Munzert (2011) in using a conditionally autoregressive (CAR) distribution where

$$\phi_j|\phi_{j\prime} \sim N\left(\frac{\sum_{j\prime \neq j} \omega_{jj\prime}\phi_{j\prime}}{\sum_{j\prime \neq j} \omega_{jj\prime}}, \frac{\sigma^2_\phi}{\sum_{j\prime \neq j} \omega_{jj\prime}}\right). \tag{2}$$

Here, the expected value of ϕ_j is the un-weighted average of $\phi_{j\prime}$ across all j's neighbours. As the variance parameter σ^2_ϕ decreases (and as the number of neighbouring constituencies increases), values of ϕ_j are smoothed more toward the average value across j's neighbours.

The non-spatially correlated random effects $\upsilon_j^{constituency}$ are modelled as

$$\upsilon_j^{constituency} \sim N(X_j\beta + \delta_{region[j]}, \sigma^2_\upsilon) \quad \text{for } j = 1, \ldots, J \tag{3}$$

where X_j gives the values of the constituency-level predictors for constituency j, β is a vector of coefficients, and δ_{region} is a random effect for the government office region in which constituency j is located.[2]

[1] α^0 is assigned a flat prior (using the `dflat()` function in BUGS). For α^k where $L_k = 2$, we set $\alpha^k_{l_1} = 0$ and assume $\alpha^k_{l_2} \sim N(0, 100)$. For α^k where $L_k > 2$, the α^k terms are modelled as draws from a common normal distribution with mean zero and standard deviation σ_{α^k}, where $\sigma_{\alpha^k} \sim Unif(0, 2)$.

[2] The β coefficients are assigned independent flat priors. The region random effects are drawn from a normal distribution with mean zero and standard deviation σ_δ, where $\sigma_\delta \sim Unif(0, 2)$

Step two: post-stratification

After estimating the model defined by equations (1)–(3), we post-stratify to get estimated constituency level opinion. The K individual-level variables in (1) define $(L_1 \times L_2 \times \ldots \times L_K) = S$ *types* of citizen, one for every possible combination of demographic characteristics. For every citizen type, indexed s, in every constituency j the estimated regression model yields a fitted probability $\hat{\pi}_{sj}$ that $y_{sj} = 1$:

$$\hat{\pi}_{sj} = \text{logit}^{-1}(\hat{\alpha}^0 + \hat{\alpha}^1_{l_1[s]} + \hat{\alpha}^2_{l_2[s]} + \ldots + \hat{\alpha}^K_{l_K[s]} + \hat{\phi}^{constituency}_j + \hat{\upsilon}^{constituency}_j). \tag{4}$$

These fitted probabilities are combined with information on the population frequency of each citizen type in each constituency, N_{sj}, to generate constituency estimates $\hat{\pi}_j$.

$$\hat{\pi}_j = \frac{\sum_s N_{sj} \hat{\pi}_s j}{\sum_s N_{sj}} \tag{5}$$

Modelling a continuous opinion variable

The above discussion assumed we are estimating constituency-level proportions for a binary opinion variable. We are also interested in political opinions measured as continuous variables, such as economic left-right ideology. In these cases, we estimate mean opinion in each constituency. Our estimation strategy for continuous opinion variables is very similar to that laid out in equations (1)–(5), except that we use a linear regression specification at the individual-level.

Estimation

All data preparation and post-stratification is performed in R (R Core Team, 2012). We estimate all multilevel regression models via Bayesian MCMC simulation using WinBUGS

(Lunn et al., 2000), with the GeoBUGS add-on for any models including local smoothing (Thomas et al., 2004). For each model, we run three separate chains of length 60,000 iterations each, the first 10,000 of which are discarded as burnin. We thin the resulting chain by a factor of 100.

Our estimation procedure thus yields 1,500 draws from the posterior distribution of average opinion for each constituency. Our point estimate of average opinion in a constituency is the mean value of these draws. We also summarise our uncertainty about average opinion by supplying posterior 95% confidence interval bounds, which are simply the 0.025 and 0.975 quantiles of the posterior sample.

3 Opinion measures and survey data

For each of the topics for which we have generated constituency opinion estimates, we detail here the survey data used and survey items selected to measure respondent opinion.

Average views on government redistribution of income (`redist11pt`, 2014) For this topic we use data from Waves 1 and 2 of the 2015 British Election Study Combined Wave 1 and 2 Panel, fielded between February and June 2014 (Fieldhouse et al., 2014). We measure respondent preferences concerning redistribution using the survey item `redistSelf` which asks

> Some people feel that government should make much greater efforts to make people's incomes more equal. Other people feel that government should be much less concerned about how equal people's incomes are. Where would you place yourself... on this scale?

The response scale runs from 0 ("Government should try to make incomes equal") to 10 ("Government should be less concerned about equal incomes"). We treat this as a continuous scale and estimate the mean score in each constituency. For Wave 2 respondents who give a valid response to this item, we use this response for estimation. For Wave 2 respondents who do not give a valid response to this item but did give a valid response to the same item in Wave 1, we use those Wave 1 responses. For those who responded to Wave 1 but not Wave 2, we include response to the item in Wave 1 where this is valid. In total we have 29,959 observations for this model.

Support for British exit from the European Union (`euref`, 2014) For this topic we use data from Wave 2 of the 2015 British Election Study Combined Wave 1 and 2 Panel, fielded between February and June 2014 (Fieldhouse et al., 2014). We measure respondent preferences concerning redistribution using the survey item `euRefVote` which asks

> If there was a referendum on Britain's membership of the European Union, how do you think you would vote?

Response options were 'Leave the EU', 'Stay in the EU', 'I would not vote' and 'Don't know'. We drop 'I would not vote' and 'Don't know' observations and estimate the proportion of voters in each constituency who would vote to 'Leave the EU'. We use observations on Wave 2 respondents, yielding 21,016 responses for this model.

Average views on the cultural impact of immigration (`immigcult`, 2014) For this topic we use data from Wave 2 of the 2015 British Election Study Combined Wave 1 and 2 Panel, fielded between February and June 2014 (Fieldhouse et al., 2014). We measure respondent views concerning the cultural impact of immigration using the survey item `immigCultural` which asks

> Do you think that immigration undermines or enriches Britain's cultural life?

Responses were recorded on a seven-point scale running from 'Undermines cultural life' (1) to 'Enriches cultural life' (7). We treat this as a continuous scale and estimate the mean score in each constituency. We use observations on Wave 2 respondents who give valid responses to the item. As a result, we have 28,096 responses for this model.

Support for same sex marriage (`ssm`, 2012-13) For this topic we used pooled data from several YouGov surveys, each of which asked the question: "would you support or oppose changing the law to allow same-sex couples to marry?" This question was asked of 7,400 respondents to YouGov polls on several dates between September 2012 and August 2013.[3] The original response format allowed respondents to indicate whether they strongly supported, tended to support, tended to oppose, or strongly opposed this change. Strong support and a tendency to support were combined to give a dichotomous variable measuring support for same-sex marriage. Don't knows were excluded from the analysis. In total we have 7,400 observations for this model.

Disapproval of Britain's membership of the EU (`eudis`, 2010) For this topic we use data from the pre-campaign wave of the BES 2010 CIPS (Clarke et al., 2014). We measure respondent EU disapproval using survey item aaq103: "Overall, do you approve or disapprove of Britain's membership in the European Union?". The response options were "strongly disapprove", "disapprove", "neither approve nor disapprove", "approve", and "strongly appove". From this we code a binary response variable equal to 1 if a respondent either "disapproved" or "strongly disapproved", and zero otherwise. In total we have 15,885 observations for this model.

Constituency ideal points on a left-right scale (`econlr`, 2010-11) For this topic we combine information from multiple survey items included in two BES surveys: the post-election wave of the BES 2010 CIPS (Clarke et al., 2014) and the post-referendum wave of the BES Alternative Vote Referendum Survey (AVRS) (Clarke et al., 2011), which contains

[3] We are grateful to YouGov for providing this data.

responses from a large number of the original CIPS respondents. Across the CIPS and AVRS surveys, we identified nine economic policy-related items with up to five response options (respondents were generally asked whether they strongly approved, approved, neither approved nor disapproved, disapproved, or strongly disapproved of a policy proposal). The nine economic policy-related survey items are detailed in Table 1.

Overall, we observe data on 10,821 individuals who responded to both the CIPS post-election wave and the AVRS post-referendum wave. Based on their answers to the nine economic policy-related items, we estimate an ordinal item response theory (IRT) model. This yields an estimate of each respondent's position on a continuous underlying left-right economic dimension. We use this as our response variable. The details of the nine policy items and of the ordinal IRT model are reported in Hanretty et al. (2014b).

4 Individual-level predictors and post-stratification

Post-stratification variables

These are the variables that are included at the individual level in (1) and which are used for post-stratification. We use data from three different sources: univariate census statistics, a sample of anonymised records (SARS) from the Census, and the survey data from which our measure of opinion comes from. Because we use multiple sources of data, we must make sure that our variable categories are comparable. Here, we describe the categories used for each of our seven post-stratification variables, and the operations necessary to reconcile different categorisations.

Gender The coding of *gender* is dichotomous (male/female); this is the same across all data sources.

Table 1: BES items used for economic left-right scaling

Item	Wording	Response options	Source
taxexempt	Exempt the first £10,000 of earnings from income tax.	1 = 'Disagree strongly', 2 = 'Disagree', 3 = 'Neither agree nor disagree', 4 = 'Agree', 5 = 'Strongly agree'	CIPS
taxmansion	Charge a 'mansion' tax on properties worth over £2 million.	1 = 'Disagree strongly', 2 = 'Disagree', 3 = 'Neither agree nor disagree', 4 = 'Agree', 5 = 'Strongly agree'	CIPS
ecotax	Introduce new eco taxes including a fuel tax for airline flights.	1 = 'Disagree strongly', 2 = 'Disagree', 3 = 'Neither agree nor disagree', 4 = 'Agree', 5 = 'Strongly agree'	CIPS
limitpension	Limit tax relief on pensions to the basic rate of tax.	1 = 'Disagree strongly', 2 = 'Disagree', 3 = 'Neither agree nor disagree', 4 = 'Agree', 5 = 'Strongly agree'	CIPS
cuts1	The Government's cuts in public expenditure are essential for the long-term health of the UK economy	1 = 'Disagree strongly', 2 = 'Disagree', 3 = 'Neither agree nor disagree', 4 = 'Agree', 5 = 'Strongly agree'	AVRS
eutax	The European Union's authority to demand tax money from the UK should be abolished.	1 = 'Disagree strongly', 2 = 'Disagree', 3 = 'Neither agree nor disagree', 4 = 'Agree', 5 = 'Strongly agree'	AVRS
tuitionfees	The Coalition Government is withdrawing all financial support to universities for teaching their students, with the exception of science subjects. University managers think this will require them to raise students tuition fees from £3,250 up to as much as £9,000. Do you approve or disapprove of this policy?	1 = 'Disapprove strongly', 2 = 'Disapprove', 3 = 'Neither approve nor disapprove', 4 = 'Approve', 5 = 'Strongly approve'	AVRS
deficittax	Do you think that tax increases are needed to help reduce the deficit?	1 = 'Big tax increases are needed', 2 = 'Small tax increases are needed', 3 = 'No tax increases are needed'	AVRS
govservices	Which of the following statements comes closest to your view?	1 = 'Good public services can only be provided by the government', 2 = 'The government should do less to provide publicly funded services and do more to encourage people to provide services for themselves.'	AVRS

Age The coding of *age* differs across our sources of data. The SARS data uses the following categories: 16-19, 20-24, 25-29, 30-44, 45-59, 60-64, 65-69, 70-74, and 75+. All other sources record age as a continuous variable. Consequently, we adopted the SARS categories. For the purposes of post-stratification, we created an artificial 18-19 age category by (1) taking the 16-19 category, and multiplying by one-quarter, to create one artificial year; (2) taking the 20-24 category, and multiplying by one-fifth, to create one artificial year; (3) adding the sum of these two categories, and using this value. We are therefore assuming that the joint distributions involving 16 to 19 year olds are very similar to the joint distributions involving 18 to 19 year olds.

Education The coding of *education* is the most problematic. We adopt the following categories, which are used in the SARS data, but which are not used in the considerably more detailed univariate statistics and survey data:

1. Qualifications data missing
2. No qualifications
3. Level 1
4. Level 2
5. Level 3
6. Level 4/5
7. Other qualifications/level unknown

These levels are similar to International Standard Classification of Education (ISCED) levels. As such, Level 4/5 corresponds to post-secondary educational attainment; Level 3 to attainment at the end of secondary education, and Levels 1 and 2 to lower secondary or primary educational attainment. Specific educational outcomes were recoded on this basis.

Marital Status The coding of *marital status* involves collapsing detailed information from SARS and from opinion data to the following dichotomy, for which information is available in the census univariate statistics:

- Married or re-married
- Single (never married), separated, divorced or widowed

Housing status The coding of *housing status* involves collapsing detailed information from the SARS and univariate census data to the following dichotomy, for which information is available in the public opinion survey data:

- Owns accommodation
- Rents accommodation

Social grade The coding of *social grade* relies on the National Readership Survey/Market Research Society social grades

1. Approximated social grade AB
2. Approximated social grade C1
3. Approximated social grade C2
4. Approximated social grade DE

Note that this refers to the social grade of the 'head of household' or 'household reference person' (HRP). For public opinion survey data, we have been able to recode information on occupation to the above categories, using (where appropriate) information on the occupation of the respondent's partner, or information on their student status. Note that this variable is not used for ILPP whenever we estimate constituency opinion using 2015 BES data, because the 2015 BES data does not currently record respondent social grade.

Private sector occupation Finally, the coding of *private sector occupation* involves a simple dichotomy between those.

- currently in private sector employment
- in public or voluntary sector employment, or unemployed

All of these seven post-stratification variables are employed for any opinion topic where we use the BES 2010 CIPS as our national survey sample. However, when we estimate opinion on same sex marriage we use YouGov polling data which does not contain information on all seven post-stratification variables. As a result, we use only five post-stratification variables in this one case. These are *gender*, *age*, *marital status*, *education* and *social grade*.

Post-stratification weights

No official UK Census information provides the constituency level joint distribution of our seven post-stratification variables. We therefore estimate post-stratification weights by combining two sources of census information. We use the Census Sample of Anonymised Records (SAR) (Office for National Statistics and Census Division and University of Manchester Cathie Marsh Centre for Census and Survey Research, 2013) which provides individual-level 2001 census responses for an anonymised sample of 5% of the population to generate estimates of the *national-level* joint population distribution of the six variables of interest.[4] Second, for each constituency j we 'rake' the national-level joint population distribution toward the *constituency-level* marginal distributions of the six variables of interest, where the latter marginal distributions are obtained from Nomis Census Area Statistics. For our constituency opinion estimates based on 2015 BES data, we use post-stratification weights generated by raking to margins which are based on univariate constituency-level results from

[4] We use Small Area Microdata from the 2001 census because, at the time of writing, the equivalent data for the 2011 census has not yet been published.

the 2011 Census (Office for National Statistics, 2011; National Records of Scotland, 2011) downloaded from NOMIS. For prior constitunecy opinion estimates using survey data from prior to 2014, we used post-stratification weights generated by raking to margins which are based on univariate constituency-level results from the 2001 Census (Office for National Statistics, 2001; National Records of Scotland, 2001) downloaded from NOMIS. (this was all that was available when we initially started the project). The result is an estimate of the joint distribution of the six variables of interest in every constituency j. The details of this procedure are as follows.

We began with the SAR data, and created a six-dimensional matrix (2 genders $\times$ 9 age categories $\times$ 7 education categories $\times$ 2 marital statuses $\times$ 2 housing statuses $\times$ 4 social grades $\times$ two sectors of the economy (private and public)). Due to the changes in the education systems of England, Wales and Scotland over time, information on the educational attainment of over 75s was not included. We therefore estimated, using those respondents in the 65-74 age group only, a multinomial model of educational attainment using all of the remaining variables in our matrix as predictors. We used the predicted probabilities of attainment in each category to create estimated counts for each cell.

We then created as many copies of this six-dimensional matrix as there were Westminster Parliamentary Constituencies (WPCs), i.e. 632. Call each of these the target matrix. For each constituency, and for each variable, we multiplied the entries in the target matrix by the proportion to which they were under-represented compared to the known marginal distribution provided by the Census Dissemination Unit. Thus, for Aberdeen North, the proportion of women in the population according to univariate census statistics (51.6%) is slightly lower than the proportion of women in the SAR (51.8%); and so all cells in the target matrix involving women were multiplied by 0.995, and all cells in the target matrix involving men were multiplied by 1.005. We finished this iterative 'raking' process when the mean absolute logged difference in these proportions was less than 0.0001. The result was an estimate of the joint distribution of variables in each constituency based on the known national joint distribution of variables as adjusted for the over-/under-representation of certain groups in each constituency.

In order to verify that these raked estimates were reliable, we compared our 2001 Census-based estimates to the limited bivariate cross-tabulations made available at WPC level by the 2001 Census (Office for National Statistics, 2001; National Records of Scotland, 2011). Here, we use tables CAS033 (Occupation by age) and CAS113 (Occupation by educational attainment). We converted these counts to notional weights by dividing by the grand sum. We can assess the congruence between our estimates and the actual Census joint distribution by calculating the absolute difference in the weight for each cell, and averaging across constituencies. The mean absolute difference for CAS033 was 1.28%; the mean absolute difference for CAS113 was 0.24%.

5 Constituency level data

Constituency level predictors

These are the variables included in the X_j matrix in (3). We select variables that are plausibly associated with average political attitudes in a constituency.

First, we include the overtly political predictors *conprev*, *labprev* and *libprev*, measures of Conservative, Labour and Liberal Democrat vote share, respectively, at the 2010 general election. When the target opinion variable is binary, we logit-transform these variables before including them in the model.

Second, we include predictors that are demographic or geographic: *density* is the natural logarithm of the ratio of the total constituency population to constituency surface area (measured in hectares, taken from the Ordinance Survey Boundary Line Service data); *earn* is the natural logarithm of median earnings in a constituency, taken from the 2013 ONS

Annual Survey of Hours and Earnings (Office for National Statistics, 2013); *nwhite* is the percentage of population that is non-white (taken from the 2011 Census); *rchristian*, *rother* and *rrefuse* measure the percentage of the population who in the 2011 Census reported being Christian, belonging to any other religion, or refused to give a religious affiliation, respectively; finally, *region* is one of eleven government regions within which constituencies are situated.

Third, we also use information from our constituency post-stratification weights in our matrix of constituency level predictors. For each constituency we take a summary measure of the central tendency of the marginal distribution of several of the post-stratification variables in this constituency and include this in X_j: For *gender*, we include the percentage of females in each constituency; For *marital status*, we include the percentage of people in each constituency who are not married; For *housing status*, we include the percentage of people in each constituency who own accommodation; For *social grade*, we assign numeric scores to each social grade ('DE'=0; 'C2'=1; 'C1'=2; 'AB'=3) and include the mean social grade score in the constituency population; For the two remaining post-stratification variables, *age* and *education*, we include in X_j the proportion of the constituency population in extreme categories. With regard to *age*, we include the proportion of the population that is aged 16 top 24 and the proportion aged 65 and above. With regard to *education*, we include the proportion of the population that have no qualifications and proportion that have qualifications at Level 4 or above.

At the moment, we use a similar set of constituency level predictors for each opinion topic. This is partly in order to maintain a stable code base across opinion topics. In addition, models with this set of predictors have been validated extensively in Hanretty et al. (2014a). However, in future we could move toward having topic-specific models including extra constituency level predictors thought to be particularly relevant for the topic of interest.

All continuous variables in X_j except for the (logit-transformed) vote share variables are re-scaled to have mean zero and standard deviation one.

Geodata

The constituency boundary data necessary to create our adjacency matrix comes from the Ordnance Survey Boundary-Line data service (Ordnance Survey, 2012). Based on this data, we create, for each constituency, a list of contiguous constituencies.[5]

References

Clarke, H., Sanders, D., Stewart, M., and Whiteley, P. (2011). British Election Study Alternative Vote Referendum Survey Data [computer file]. Downloaded from http://www.bes2009-10.org/av-ref.php.

Clarke, H., Sanders, D., Stewart, M., and Whiteley, P. (2014). British Election Study, 2010: Campaign Internet Data [computer file]. Colchester, Essex: UK Data Archive [distributor]. SN 7530, http://dx.doi.org/10.5255/UKDA-SN-7530-1.

Fieldhouse, E., Green, J., Evans, G., Schmitt, H., and van der Eijk C. (2014). British Election Study Internet Panel Waves 1 & 2. Downloaded from http://www.britishelectionstudy.com/data-object/2015-british-election-study-combined-wave-1-and-2-panel/.

Hanretty, C., Lauderdale, B., and Vivyan, N. (2014a). Comparing Strategies for Estimating Constituency Opinion from National Survey Samples. Working paper, Durham University.

Hanretty, C., Lauderdale, B., and Vivyan, N. (2014b). Dyadic Representation in a Westmin-

[5] Five UK constituencies have no (contiguous) neighbours in the data. We manually assign each of these islands a single 'neighbour', the nearest mainland constituency, where 'nearest' means smallest point-to-point distance.

ster System. Presented at the Anual Meeting of the European Political Science Association, Edinburgh, UK.

Lunn, D., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10.:325–337.

National Records of Scotland (2001). 2001 Census: Aggregate data (Scotland) [computer file]. UK Data Service Census Support. Downloaded from http://www.scrol.gov.uk/. This information is licensed under the terms of the Open Government Licence [http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2].

National Records of Scotland (2011). 2011 Census: Aggregate data (Scotland) [computer file]. UK Data Service Census Support. Downloaded from http://www.scrol.gov.uk/. This information is licensed under the terms of the Open Government Licence [http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2].

Office for National Statistics (2001). 2001 Census: Aggregate data (England and Wales) [computer file]. UK Data Service Census Support. Downloaded from: https://www.nomisweb.co.uk/. This information is licensed under the terms of the Open Government Licence [http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2].

Office for National Statistics (2011). 2011 Census: Aggregate data (England and Wales) [computer file]. UK Data Service Census Support. Downloaded from https://www.nomisweb.co.uk/. This information is licensed under the terms of the Open Government Licence [http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2].

Office for National Statistics (2013). 2013 Annual Survey of Hours and Earnings: Constituency Level Estimates. Downloaded from `https://www.nomisweb.co.uk/`.

Office for National Statistics and Census Division and University of Manchester Cathie Marsh Centre for Census and Survey Research (2013). Census 2001: Individual Licenced Sample of Anonymised Records (I-SAR) [computer file]. Colchester, Essex: UK Data Archive [distributor]. SN: 7205, `http://dx.doi.org/10.5255/UKDA-SN-7205-1`.

Ordnance Survey (2012). OS Boundary Line Westminster Constituencies [Shapefile geospatial data]. OS Boundary-Line Data Service. Downloaded from: `http://www.ordnancesurvey.co.uk/oswebsite/products/boundary-line/index.html`.

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Selb, P. and Munzert, S. (2011). Estimating constituency preferences from sparse survey data using auxiliary geographic information. *Political Analysis*, 19(4):455–470.

Thomas, A., Best, N., Lunn, D., Arnold, R., and Spiegelhalter, D. (2004). *GeoBUGS User Manual Version 1.2*.