JOURNAL OF SPATIAL INFORMATION SCIENCE
Number 16 (2018), pp. 117–136

doi:10.5311/JOSIS.2018.16.377

RESEARCH ARTICLE

Locally-varying explanations behind the United Kingdom's vote to leave the European Union

Roger Beecham[1], Aidan Slingsby[2], and Chris Brunsdon[3]

[1]University of Leeds, UK
[2]City, University of London, UK
[3]Maynooth University, Republic of Ireland

Received: August 11, 2017; returned: October 19, 2017; revised: October 30, 2017; accepted: November 11, 2017.

Abstract: Explanations behind area-based (Local Authority-level) voting preference in the 2016 referendum on membership of the European Union are explored using aggregate-level data. Developing local models, special attention is paid to whether variables explain the vote equally well across the country. Variables describing the post-industrial and economic "successfulness" of Local Authorities most strongly discriminate variation in the vote. To a lesser extent this is the case for variables linked to "metropolitan" and "big city" contexts, which assist the Remain vote, those that distinguish more traditional and "nativist" values, assisting Leave, and those loosely describing material outcomes, again reinforcing Leave. Whilst variables describing economic competitiveness co-vary with voting preference equally well across the country, the importance of secondary variables—those distinguishing metropolitan settings, values and outcomes—does vary by region. For certain variables and in certain areas, the direction of effect on voting preference reverses. For example, whilst levels of European Union migration mostly assist the Remain vote, in parts of the country the opposite effect is observed.

Keywords: European Union, referendum, multi-level modelling, geographically-weighted statistics, LASSO, area-based analysis

1 Introduction

On 23rd June 2016 the United Kingdom (UK) held a referendum, for the second time, on its political association with Europe. The outcome of the first, on 5th June 1975, was emphatic, with 67% voting for continued membership of the then European Community. The result of the 2016 vote, this time on membership of the European Union (EU), was more fractured. Whilst the overall outcome was a slight preference for leaving the EU (51.9%), the result varied greatly across the country. Of the 11 regions of Great Britain (GB), only Scotland and London voted in favor of Remain (62% and 60% respectively).

Political Scientists have argued that the spatially-fragmented vote is symptomatic of widening social divisions in the UK linked to structural change [11]. Since the 1970s, opportunities have increasingly concentrated in the South East of England and particularly London, whereas more provincial locations, particularly those cities and regions with economies once monopolized by single-industries, saw limited growth and depopulation [6]. The social-geography of the UK is, then, today more socially distinctive than it was in 1975. Thus, spatial differences in the 2016 referendum vote might be linked to place-based factors: Local Authorities recording the strongest preference for Leave have been described as "left-behind" places characterized by chronic low skills, socially conservative and nativist values, whilst those associated strongly with Remain with more affluent, highly-educated and diverse populations [10].

This study follows that of earlier area-based analyses [11,12,16] in examining the extent to which Local Authority-level (LA) differences in voting preference can be accounted for by differences in the demographic composition of LAs: that is, it considers how reflective the narrative of the "left-behind" is of reality. A unique flavor to this analysis is its attention to local variation. We study whether modelled explanations hold equally well across the country. After exploring the geography of the referendum vote with global regression models, we develop local multivariate models and compare differences in the covariates suggested by these local models—their effect size, direction and the "concepts" they represent. We further explore and depict how associations between candidate explanations vary spatially using geographically-weighted statistics [3]. Finally, we attempt to qualify the scale at which different variables operate using multi-level modelling. We find that:

- As in earlier studies [11, 12, 16], differences in LA-level voting preference are most strongly associated (negatively) with education outcomes, particularly when expressed by degree-level education. Our study adds that *degree-educated* stands apart as the obvious covariate in that its strong negative association with Leave is true across GB. That we observe a similar pattern of association with variables describing occupation structure, specifically those working in *professional occupations*, supports the reading that the geography of Leave is structured around those places economically "left behind" after de-industrialization.
- A smaller effect is contributed by variables distinguishing metropolitan and "big city" contexts from more provincial town and rural settings.
- Variables that distinguish social values and cultural diversity have an additional effect. After controlling for differences in education levels, variables that describe traditional and nativist values (*Christian* and ethnic diversity) serve to assist the LA-level share of Leave vote. It is in certain of these variables, as well as those differentiating metropolitan contexts, where effects are more locally specific.

2 Background

In the months after the 2016 referendum on membership of the EU, numerous area-based analyses appeared online and in-print (e.g., [4, 11, 12, 16, 20]) using results data published at LA-level by the Electoral Commission[1].

Goodwin & Heath [11] compared the results data against 2011 Census variables describing the the age, ethnicity, migration structure and education outcomes of residents in LAs, along with indicators of historical euroscepticism. The authors developed multivariate models, with LA share of Leave vote as the outcome, and found differences in education levels between LAs to have the largest effect in explaining differences in the Leave vote; that to a lesser extent the age structure and history of migration have an effect; and that there is substantial residual variation in Scotland, where the strength of support for Leave is much lower than that modelled using demographics alone.

Harris & Charlton [12] took a similar approach, but used a multi-level framework (LAs within regions) to distinguish variation operating at the region-level from variation that is within-region. Similar to Goodwin & Heath [11], they highlight the effects of age and occupational structure of LAs in explaining variation in the vote as well indicators describing levels of immigration. The authors also attend to spatial patterning of model residuals given the known geography of the Leave vote. Comparing their multivariate model for England & Wales against a *null* model (intercept 0, slope 1) they found that, net of differences in the demographic composition of regions, the majority of unexplained variance remains between regions: especially for London, but also West Midlands, East, South East and East Midlands, their model underestimates the Leave vote; the corollary is true of North East, North West and Wales. This direction of regional variation is opposite to that reported by Goodwin & Heath [11]. When including Scotland and London as dummy variables, their model overestimates the Leave vote for LAs in London. The alternative finding might be explained by differences in method and the fact that Harris & Charlton's [12] model excludes Scotland. It might also relate to the way in which the two studies defined the outcome variable. Goodwin & Heath expressed support for Leave as a proportion of the total vote in a LA, whereas Harris & Charlton's outcome was an odds ratio taking into account two quantities: *observed* support for Leave, defined as each LA's Leave vote as a share of the total Leave vote, and the *expected* vote, defined as each LA's size of electorate as a share of the total electorate. Note that Harris & Charlton's outcome variable can be considered as a broader indicator of support for Leave than that of Goodwin & Heath—it measures those actively voting to support Leave from the population that had *opportunity* to vote. Since it is conflated with turnout, however, this broader measure could of course increase irrespective of the *relative* support for Leave | Remain.

A further paper considering the geography of the EU referendum vote was published by Manley et al. [16]. The authors used individual-level polling data collected over the year prior to the referendum vote (British Election Study (BES)[2], March 2015–March 2016) to estimate support for Leave across the 380 LAs in GB. They did so using a multi-level framework that considers individual respondents from the polling data within LAs. Such a formulation allows additional variation, or "distinctiveness," at the LA-level to be quantified after accounting for age and qualification differences in respondents. Manley et al. describe a matrix of probabilities of outcome for each LA in GB, which was then scaled ac-

[1]https://www.electoralcommission.org.uk/
[2]http://www.britishelectionstudy.com

cording to the demographic composition of LAs, as defined by 2011 Census. Not all groups are equally likely to vote, and a final weighting was applied assuming equivalent relative turnouts by age and qualification groups to the 2015 general election. The expected counts generated from this modelling process were compared against observed results and the distribution of residuals analysed. Manley et al. made the observation that the pattern of turnout and voter registrations was different in the referendum than in recent general elections. Turnout was much higher and there was an increase in voter registrations in certain parts of the country immediately prior to the vote. Since there is a strong association between turnout and age it was assumed that higher levels of LA-level turnout indicate a greater relative share of the vote from younger residents. Increases in voter registration prior to the referendum might also suggest greater interest amongst younger voters. An enlarged youth vote is likely to bolster Remain and when included in Manley et al.'s model, effects were observed in this direction: the Leave vote was lower than expected in places with high turnout and increased registration in the months prior to the referendum. On the subject of spatial patterning (region-level) in residuals, Manley et al. found only slight additional variation due to region for London and Scotland and to a lesser extent the North West. In these locations the observed Leave vote was smaller than their model expected.

Despite differing slightly in method and interpretation, a consistent set of themes was identified in these three papers. LA-level variation in support for Leave can be accounted for by differences in the educational and occupational outcomes, age and to a lesser extent immigration profile of LAs, with some evidence of unexplained regional and more idiosyncratic LA-level effects. Given these modelled relationships, and the obvious geographic differences in the vote (Figure 2), the popular discourse around the Leave vote representing that of the "left-behind" [5, 8] gains legitimacy. Not directly considered in these studies, though, is whether the explanations hold equally well for different parts of the country. For example, do levels of education explain variation equally as well in Scotland as in the South West? Might secondary variables, such as those describing EU-born migration, be more relevant in parts of the country experiencing the greatest change due to that migration? The notion of locally-varying explanation forms the basis of this study.

3 Data and methods

Referendum data describing the number of votes for Leave, the number of votes for Remain, as well as the size of electorate have been released by the Electoral Commission[3]. We combined this count dataset of voter behavior with demographic and socio-economic data measured through the 2011 Census, aggregating to LA-level.

Population characteristics were selected based on the media discourse around the Leave vote: that of the "left-behind" and of the varying experiences of de-industrialization [5]. These variables are listed in Table 1. The proportion of *degree-educated* residents and residents working in *professional* occupations appear particularly prescient given the media discourse around the Leave vote and post-industrialization: highly-educated, professionalized workers epitomize the so-called *knowledge-economy* [7]. The proportion of younger adults, a generally more upwardly mobile group, might also provide some indication of a LA's economic attractiveness and level of opportunity. So too might variables associated with ethnic and cultural diversity. The *no car household* and mode of *travel to work* variables

[3]https://www.electoralcommission.org.uk/

might initially appear to be a strange choice. Car ownership has traditionally been used in social research as *the* proxy for household income. Whilst this may still be true of certain parts of the country, when analysed at LA-level it is also likely to discriminate metropolitan, "big city"-type living from those living in satellite towns and more rural areas—a further apparent dividing line of the Leave | Remain vote. Associations between the selected variables listed in Table 1 and the Leave vote are presented graphically in Figure 1.

variable	notes	justification/theory
Degree-educated		post-industrial / knowledge-economy
Professional occupations		
Younger adults	Age 20-44	
English speaking	Speak very well / main language	diversity / values / outcomes
Single-ethnicity	Household level	
Not good health	Fair/bad/very bad health	
White British/Irish		
Christian		
EU-born (not UK)		
Own home	Household level	metropolitan / "big city"
Don't own car	Household level	
Private transport to work		

Table 1: Proposed 2011 Census variables for explaining Local Authority share of Leave vote.

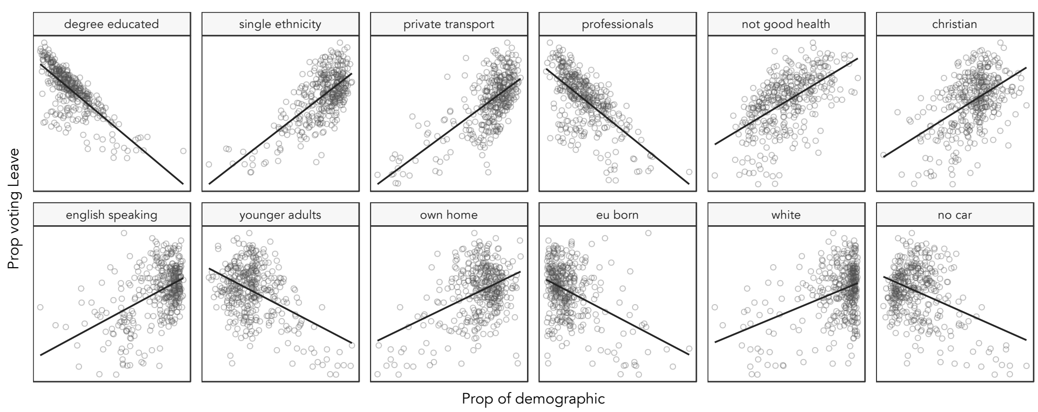


Figure 1: Scatterplots of candidate 2011 Census variables against the outcome variable—share of Leave vote by LA. Variables are ordered left-to-right, top-to-bottom by *Pearson's* correlation coefficient and are annotated with linear regression lines fit via ordinary least squares.

The effect of candidate explanatory variables on the Leave vote was investigated more formally using a regression framework. For the multivariate models presented in sections 4.2 and 4.3, variable selection and regularization was performed using the least absolute shrinkage and selection operation (LASSO) [22]. Since our research questions are concerned with possible spatial variation in factors explaining LA-level voting preference, we investigated how well the models hold for different parts of the country. We did this by attending to the spatial distribution of residuals from our global models and using geographically-weighted summary statistics [3]. Finally, we attempted to establish the scale at which different explanatory variables operate by generating multi-level models (LAs within region) for each explanatory variable and identifying the level of variance explained between regions and between LAs within regions with the addition of each variable.

Note that this analysis covers the 380 LAs of Great Britain only. Not included is Northern Ireland (56% Remain) and Gibraltar (96% Remain). Gibraltar is clearly a very separate polity. One reason for excluding Northern Ireland is that result data are available only at the region level and not disaggregated by LA. Another, given this study's focus on local and *spatially* varying explanation, is that both are physically disconnected from mainland Great Britain. Were more detailed data to become available, analysis and comparison of Scotland and Northern Ireland would be instructive, especially so since the Leave vote is so heavily concentrated in England & Wales (e.g., Figure 2).

Data processing and analysis was carried out within the `R` statistical programming environment. The LASSO procedure was performed using the `glmnet` package [19], geographically-weighted statistics using `GWmodel` [15] and multi-level modelling using `lme4` [2]. Measures of uncertainty (confidence intervals) around regression parameter estimates were estimated through a bootstrap. Full data analysis scripts and explanation can be accessed at this github repository: `github.com/rogerbeecham/brexit-analysis/`.

4 Analysis

4.1 Characterising LA-level variation in voting preference: comparison against a uniform model

We start by comparing the results data against a uniform model that assumes each LA votes for Leave in the same proportion as Great Britain (GB) as a whole. For continuity with later sections, we can express this as an intercept-only model where y_i is the Leave vote in each local authority, $\beta_1 = 0$ and $\beta_0 = \bar{y}$ (Leave vote for GB as a whole).

$$y_i = \beta_0 + \beta_1 + \varepsilon_i$$

Plotting residuals from this uniform model results in maps that were widely published in the aftermath of the vote (Figure 2). The uniform model overestimates the Leave vote in Scotland and London, whilst the corollary is true for much of England and Wales. There are some noticeable local exceptions: the dark red of Cambridge representing a much smaller Leave vote (26% of the vote share) than the GB average (52%). Comparison between a conventional Choropleth and equal-area cartogram in Figure 2 emphasizes the metropolitan patterning of the Remain vote. So too does the asymmetric shape to the legend, which presents LAs in ascending order according to the relative size of the Leave vote. Notice the wide bars at the tail of pro-Remain LAs (in red), suggesting the most pro-Remain LAs were more emphatic than the most pro-Leave LAs: the share of Remain for the top 10 most pro-Remain LAs was 76%, whilst the equivalent value for the most pro-Leave LAs was 72%. That the very obvious regional patterns in voting behavior exist—the sharp contrast with Scottish and London LAs in particular—we can ask, as do others [11, 12, 16], whether there is something fundamentally different about LAs in different parts of the country that gives rise to this pattern of variation.

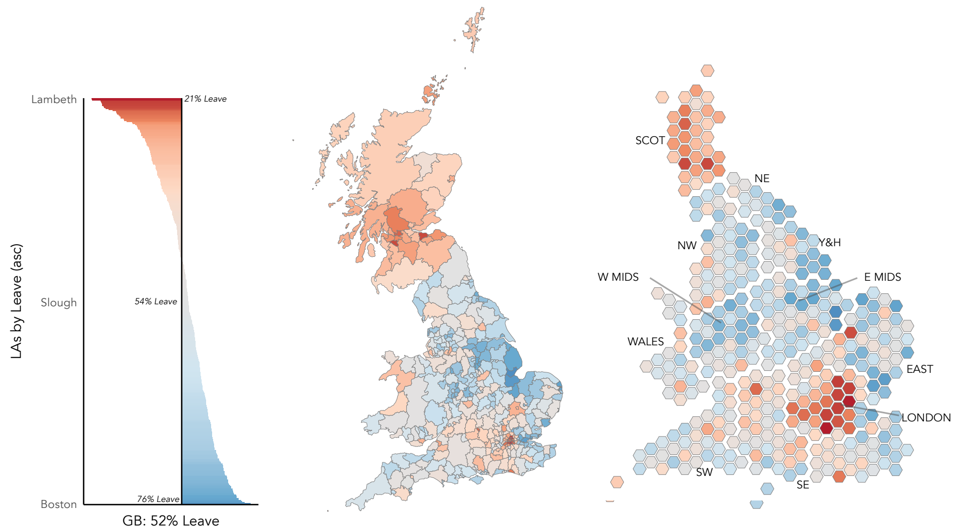
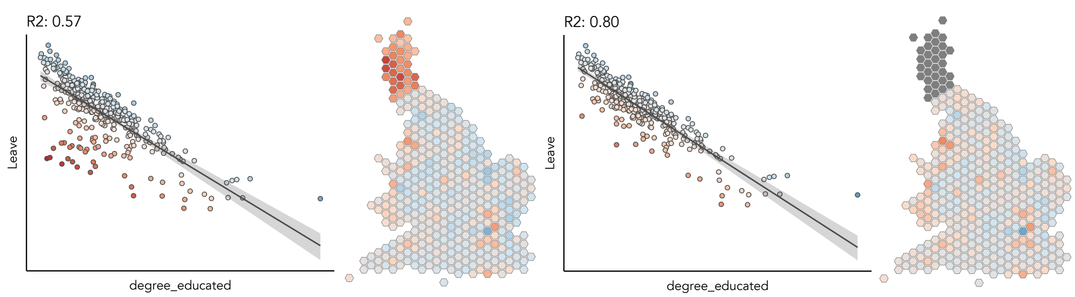


Figure 2: Residuals from a uniform model that assumes each LA votes for Leave in the same proportion as GB. Pro-Leave LAs (relative to the GB average) are blue and pro-Remain are red. A ColorBrewer RdBu diverging scheme [13] is used. LA outlines contain National Statistics data © and OS data © Crown copyright and database right 2015. Equal-area cartogram outline adapted from Ben Flanagan, ESRI UK.

4.2 Explaining LA-level variation in the vote using global models



(a) (b)

Figure 3: LAs colored by residuals from univariate models regressing Leave on *degree-educated*. The scatterplots are accompanied with regression lines and 95% confidence intervals derived from a bootstrap. 3a) model fit to whole of GB; 3b) model fit to England & Wales only. Equal-area cartogram outline by Ben Flanagan, ESRI UK.

To investigate the extent to which differences in the vote relate to differences in local, area-based characteristics, we replace our uniform model with a standard linear regression such that the Leave vote is expressed as a linear function of the *degree-educated* variable—that is, the proportion of residents in a LA educated to bachelors level or higher (d_{i1}):

$$y_i = \beta_0 + d_{i1}\beta_1 + \varepsilon_i$$

Presented in Figure 3 is a summary of residuals resulting from this model. *Degree educated* is the most heavily correlated variable (inversely) with Leave and explains 57% of the variation in voting preference between LAs in GB. A cursory glance at the the scatterplot in Figure 3 reveals that residuals are not evenly distributed around the regression line: there is a group of LAs where the univariate model overestimates the Leave vote given the levels of education in those areas and the accompanying map shows these LAs to be overwhelmingly concentrated in Scotland. By excluding LAs in Scotland, the residuals distribute more randomly around the regression line and 80% of variation in LA-level voting preference in England & Wales (EW) can be explained with variation in the *degree-educated* variable. There nevertheless remains spatial structure to the residuals: they do not distribute randomly around the country (Moran's I 0.76 GB model; 0.43 EW model).

As demonstrated by the scatterplots in Figure 1, other variables are associated with Leave and might be used to explain additional variation net of education levels. We investigate the effect of these variables by building a global multivariate regression model. This model is fit using the least absolute shrinkage and selection operation (LASSO) [22]. LASSO is a procedure for model regularization and selection that loosely follows Occam's razor heuristic: among competing solutions that fit the data equally well, the one with the fewest assumptions—the fewest variables in this case—is selected. LASSO is similar to ordinary least squares in that for any model specification it tries to minimize the sum of residuals in the same way; however it also adds an upper bound (t) on the sum of regression coefficients since large coefficients are likely to lead to prediction error and overfitting. An advantage of the LASSO procedure over other forms of regularization is in variable selection: the least important coefficients are shrunk to zero, effectively removing those covariates from the model.

$$y_i = \beta_0 + x_{i1}\beta_1 + ... + x_{ik}\beta_k + \varepsilon_i$$

$$\text{in matrix form:} \quad Y = X\beta + \varepsilon$$

ensuring that:

$$\text{minimize} \left(\frac{||Y - X\beta||_2^2}{n} \right) \text{subject to} \sum_{j=1}^{k} ||\beta||1 \leq t$$

Since LASSO imposes constraints on the size of estimated coefficients, it is necessary to standardize explanatory variables before running the procedure. This process is handled by the `glmnet` package [19] and for interpretation coefficients are returned in the original units on which they are based.

Presented to the left of Figure 4 are coefficients from a multivariate model fit using LASSO. All 12 variables in Figure 1 were passed to this model; also added were dummy variables distinguishing LAs that are within London and Scotland. The line through the regression coefficients in Figure 4 and their transparency is determined by 95% confidence intervals calculated via a bootstrap.

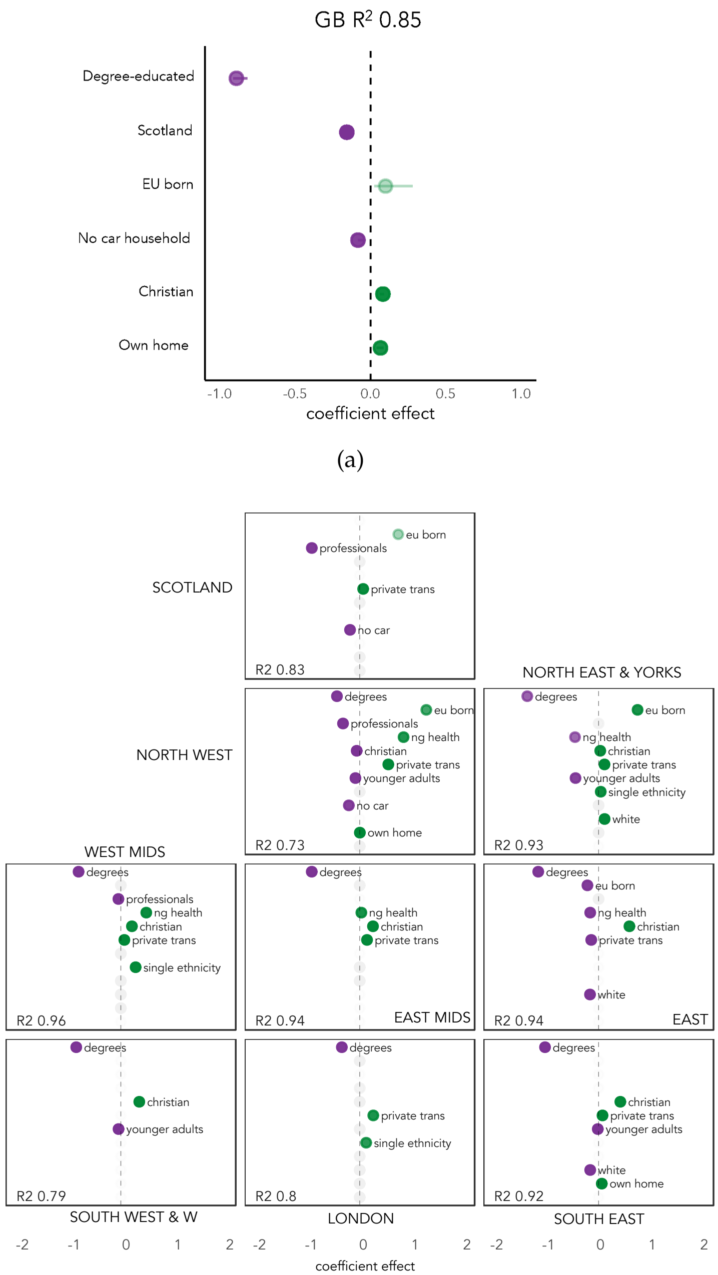
The model created under this LASSO procedure identified six variables. *Degree-educated* contributes the largest coefficient effect. Holding the other variables constant, a one percent point increase in the *degree-educated* population decreases the leave vote by 0.9 percent points. The fact that Scotland is selected by the LASSO procedure is instructive: there is something fundamentally different about Scotland, not accounted for completely by census variables, that lowers preference for Leave (by 16% points after controlling for demographics). The effect of the *EU-born* variable is counter to that expected. In Figure 1 the variable appears negatively correlated with Leave and we speculate might represent economic opportunity and relative diversity. After controlling for variation in other demographic characteristics, the model suggests an increase in the *EU-born* population in fact *increases* the Leave vote. Notice, however, the large confidence interval around this coefficient. Given the resampling procedure used to generate our bootstrap, this interval indicates that the effect of *EU-born* is likely to vary across LAs.

4.3 Region-specific explanations implied by local models

Whilst the inclusion of more than one explanatory variable clearly improves model fit—we see an improvement for GB as a whole from $0.57 \rightarrow 0.85$ (adjusted R^2)—this improvement does not preclude the possibility that certain variables better explain variation in certain parts of the country than others. That the residuals from the multivariate model still vary systematically with space (Moran's I 0.51) is evidence to support such an assertion. To investigate the possibility of locally-varying explanation, we fit separate models for separate regions of GB. Doing so using LASSO allows selection of variables that may be uniquely discriminating given local data. Since there are just 380 LAs in GB, and the size of regions vary (there are only 12 LAs in the North East), we merged some neighboring regions to create nine super-regions: East (47 LAs), East Midlands (40 LAs), London (33 LAs), North East and Yorkshire & Humber (33 LAs), North West (39 LAs), Scotland (32 LAs), South East (67 LAs), South West & Wales (59 LAs), West Midlands (30 LAs). Regression coefficients from these local models appear in the right of Figure 4 using an identical visual mapping to that used to present the global model.

London is a region that includes many LAs posting very strong support for Remain (Lambeth 79% and Hackney 78% for Remain) as well as several pro-Leave authorities to the east (Havering 70%, Bexley 63% and Barking & Dagenham 62% for Leave). In addition to the *degree-educated* variable, which we consider a proxy for post-industrial success, the model suggests variables that distinguish metropolitan and "big city" contexts (*private transport to work* and *single ethnicity household*). The coefficient effects given by the model are in the direction implied by the scatterplots in Figure 1, but the effect contributed by *degree-educated* is smaller (-0.34) than observed in most other local models presented in Figure 4.

The East, South East and South West & Wales super-regions are similar to the extent that, after controlling for variation in *degree-educated* and those variables describing the age and ethnic structure of LAs, the *Christian* variable serves to increase the Leave vote with a reasonably large effect size—coefficient of 0.59, 0.42, and 0.35 respectively. This variable appears in the same direction for all other local models with the exception of London,



(a)

(b)

Figure 4: Coefficients for multivariate models fit to data for GB (4a) and super-regions (4b) and annotated with adjusted R^2. Positive coefficients are green, negative purple and color lightness varies according to a 95% Confidence Interval calculated via a bootstrap. Note that the GB model was specified with additional dummy variables for Scotland and London.

Scotland and the North West; it may represent a proxy for less dynamic, more "traditional" contexts and seems consistent with the popular discourse around the Leave vote. However, for the model built using LAs in the East, we observe effects that confound expectation: after controlling for variation in key variables (*degree-educated* and *Christian*), the *not good health* variable, which might relate to material disadvantage, and *white*, representing low diversity, have the effect of slightly lowering the Leave vote (coefficient of -0.16). Note that although the East region is more pro-Leave than the GB average, especially so towards the coast (Castle Point 73% and Thurrock 72% for Leave), it does contain some strong pro-Remain LAs with very different local contexts and economic histories (Cambridge 74% and St. Albans 63% for Remain).

For the East Midlands and West Midlands, in addition to *Christian* (coefficient effect of 0.25 and 0.21 respectively), variables are selected that discriminate metropolitan from more provincial locations (*private transport to work*). The *not good health* variable also appears in these models and serves to increase the Leave vote.

The least successful model at explaining variation in voter preference is in the North West region—this, despite nine variables being selected by the LASSO. Note that for this model the size of effect contributed by *degree-educated* is comparatively small and that, as with the two regions of the Midlands, *not good health* assists the Leave vote. For the North West, but also North East & Yorkshire and Scotland, the *EU-born* variable is given large positive coefficients (1.3, 0.75 and 0.74 respectively). Again this effect is opposite to that expected given the global pattern of correlation between *EU-born* and Leave in Figure 1. It should be noted that Figure 4 reports coefficients reprojected to their original scale—the large coefficients should be treated cautiously since for *EU-born* is a low value and range variable.

For Scotland, *professionals* is identified ahead of *degree-educated*; this makes sense since both are conceptually related and when studying simple associations (Pearson's correlation coefficient r.) between the explanatory variables in Figure 1, Scotland is unique in that *professionals* is more strongly associated with Leave ($r. - 0.76$) than *degree-educated* ($r. - 0.59$). With the exception of the *EU-born* variable, whose coefficient effect comes with high uncertainty, the model suggested for Scotland shares some conceptual similarity to London: variables associated loosely with "post-industrialization" and separating metropolitan or "big city" contexts (*private transport*, *no car*) are selected. Absent from the London, Scotland and North West models are variables representing "traditional" or nativist values (e.g., *Christian*, *white* and *single-ethnicity* household); this a departure from other regions of GB.

4.4 Exploring the scale of locally-varying explanation

In generating models aggregated to super-region level we assume that super-regions represent unique contexts and that the *relationships* between Leave and possible covariates are discrete with these contexts. Geographically-weighted statistics [3] provides a mechanism for exploring the extent to which relationships might vary *continuously* over space. Figure 5 displays geographically-weighted correlation coefficients for each population variable against the Leave vote. The coefficients are calculated separately for each LA using an adaptive bandwidth set to $n = 50$: that is, for each LA, a spatial window is created containing its 50 nearest neighbors and attribute values at each LA within this window are weighted inversely according to distance. The decision on bandwidth selection was ar-

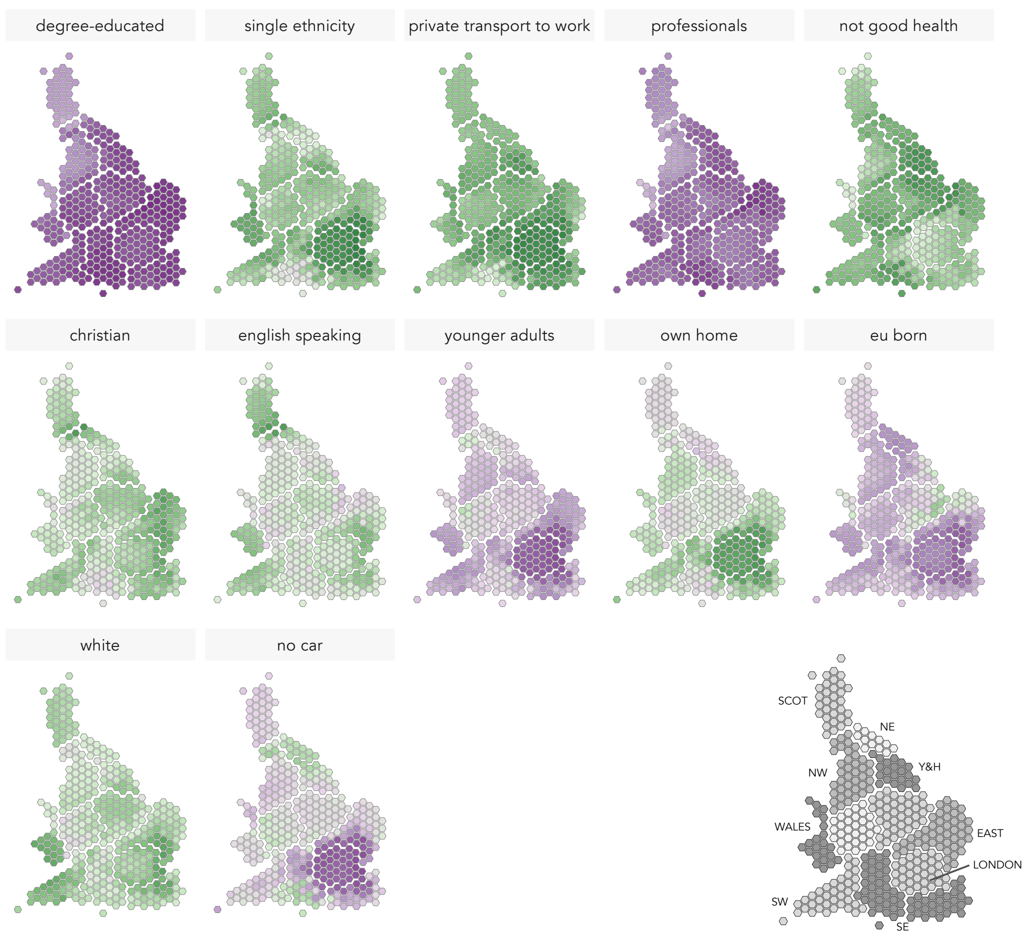


Figure 5: Geographically-weighted correlation coefficients for candidate explanatory variables against share of Leave vote. Coefficients are mapped to a $green(+ve) \rightarrow purple(-ive)$ diverging color scheme. A ColorBrewer PRGn diverging scheme [13] is used. Equal-area cartogram outline adapted from Ben Flanagan, ESRI UK.

rived at using heuristics rather than an objective function. This size of bandwidth (number of LAs) generally corresponds to that of the super-regions on which the local models were based—these range from 30 (West Midlands) to 67 (South East) LAs. Additionally, exploratory visual analysis of geographically-weighted correlation structure was performed at different bandwidths and $n = 50$ presented as a resolution sufficiently large to avoid obviously spurious structure.

Studying the maps in Figure 5, *degree-educated* and *professionals* might be regarded as "global" predictors. Although the strength of the negative association with Leave sometimes varies (for example, the North West and Scotland for *degree-educated*), the direction remains the same throughout GB. *Private transport to work* and *not good health* are both associated positively with Leave; the latter variable particularly so towards the Midlands, parts of the South East and South West and Yorkshire & Humber. A cursory glance at Figure 5 reveals London's unique influence. The *no car* variable is very strongly negatively associated with Leave around London; elsewhere this variable is less discriminating and in fact is positively associated with Leave in parts of the North East, South East and South West. A similar pattern, though in the opposite direction, is present in the *own home* variable. The association with variables related to "diversity" and "values" is interesting. *Christians* has a strong and positive relationship with Leave in the East and with LAs towards the Scottish border and to a lesser extent the South West. This spatial pattern is also true of the *English-speaking* and *white* variables. Notice also the pattern in *EU-born*. For most of the country this variable, which might be a proxy for "diversity" or economic opportunity, is negatively associated with Leave. For parts of East Anglia and Lincolnshire, bordering the East Midlands and East regions, the relationship is reversed. It should be noted here though that analysing correlation coefficients in isolation and outside of the regression framework, we cannot quantify the effect of these variables after partitioning on variation in other key variables. In the local models fit to super-regions, *EU-born* served to increase the Leave vote in Scotland, North West and North East & Yorkshire—parts of the country where, according to the geographically-weighted correlation coefficients, that variable typically has a negative association with Leave.

Finally, we can attempt to quantify the (administrative) *scale* at which our candidate explanatory variables operate using a multi-level modelling framework that considers LAs (level 1) within regions of GB (level 2). We again start with an intercept-only model, where β_0 is the overall mean proportion in favor of Leave, but where residuals are split into two levels: that between groups (the difference between the region mean and the overall mean (u_j)) and that within groups and between individuals (the departure of individual LAs from their regional means).

$$y_{ij} = \beta_{0j} + \beta_1 + \varepsilon_{ij} \qquad (\beta_1 = 0) \qquad \text{intercept-only model}$$

$$\beta_{0j} = \beta_0 + u_j \qquad \text{with group-level intercepts}$$

y_{ij}	is the estimated Leave vote for an individual LA (i) within its region (j)
β_{0j}	is the estimated intercept for a region (j)
ε_{ij}	is the error term for an individual LA (i) within its region (j)

From this intercept-only model fit to LAs in England & Wales, we find that 34% of total variance in the Leave vote can be attributed to differences between regions. This reinforces the observation that can be made graphically (Figure 2)—that the Leave vote is to an extent regionally structured, otherwise all variance would exist between LAs.

In identifying the *scale* at which different candidate explanatory variables operate, we analyse how this relative division of variation changes when accounting for local demographics. We build separate models adding each explanatory variable as a fixed effect ($\beta_{1j}x_{ij}$). In these models the effects of explanatory variables are assumed to be constant across regions, but as in the intercept-only model, the distribution of explanatory variables and outcome may change from region-to-region and we allow the intercept to vary in the same way ($\beta_0 + u_j$). We then consider the division of variance after controlling for each demographic characteristic.

$$y_{ij} = \beta_{0j} + \beta_1 x_{ij} + \varepsilon_{ij} \qquad \text{varying intercept, fixed explanatory variable}$$

Since our explanatory variables are measured at the LA-level (level 1), their addition necessarily reduces the level 1 variance. However, the level 2 variance provides some indication of the distribution of explanatory variables between regions. Including *degree-educated* as a covariate, we find that not only does the overall variance reduce substantially, but that so too does the variance between regions: after controlling for differences in levels of *degree-education*, 16% of the unexplained variance is at the region level; for the intercept-only model this figure is 34%. Such a division implies that observed vote shares for LAs are not tightly packed around their intercept-adjusted regression lines and that, when controlling for *degree-level* education, the region an LA belongs to does not have a substantial impact on the outcome (Leave). By contrast, the addition of *not good health* slightly increases the relative share of variance that is between-region (36%); thus the distribution of *not good health* is likely regionally-specific. Figure 6 supports this interpretation. Regression lines, offset by regional intercepts, are presented for models fit separately to each explanatory variable. Also presented are estimates of pseudo-R^2 for these models [17]: marginal R^2 describes the proportion of variance explained by the fixed (non-regional) factors alone, conditional R^2 is the proportion of variance explained by both the fixed and random (varying-intercept) factors.

By allowing the slope ($\beta_{1j}x_{ij}$) as well as intercept (β_{0j}) to vary by region, we can examine the extent to which relationships or *effects* of variables change by region explicitly.

$$\begin{aligned} y_{ij} &= \beta_{0j} + \beta_{1j} x_{ij} + \varepsilon_{ij} \qquad \text{varying intercept and slope} \\ \beta_{1j} &= \beta_1 + u_{1j} \end{aligned}$$

This results not only in regression lines adjusted variously according to the distribution of explanatory variables and outcome, but with slopes adjusted by regionally-specific effects. Parameters additional to the varying intercept model are the between-region variance in the slopes of each explanatory variable and the correlation between regional slopes and intercepts. Figure 7 is encoded in the same way as the random-intercept models. To test whether the effect of explanatory variables does indeed vary across regions, we also present a test statistic comparing the random slope model with the random intercept model

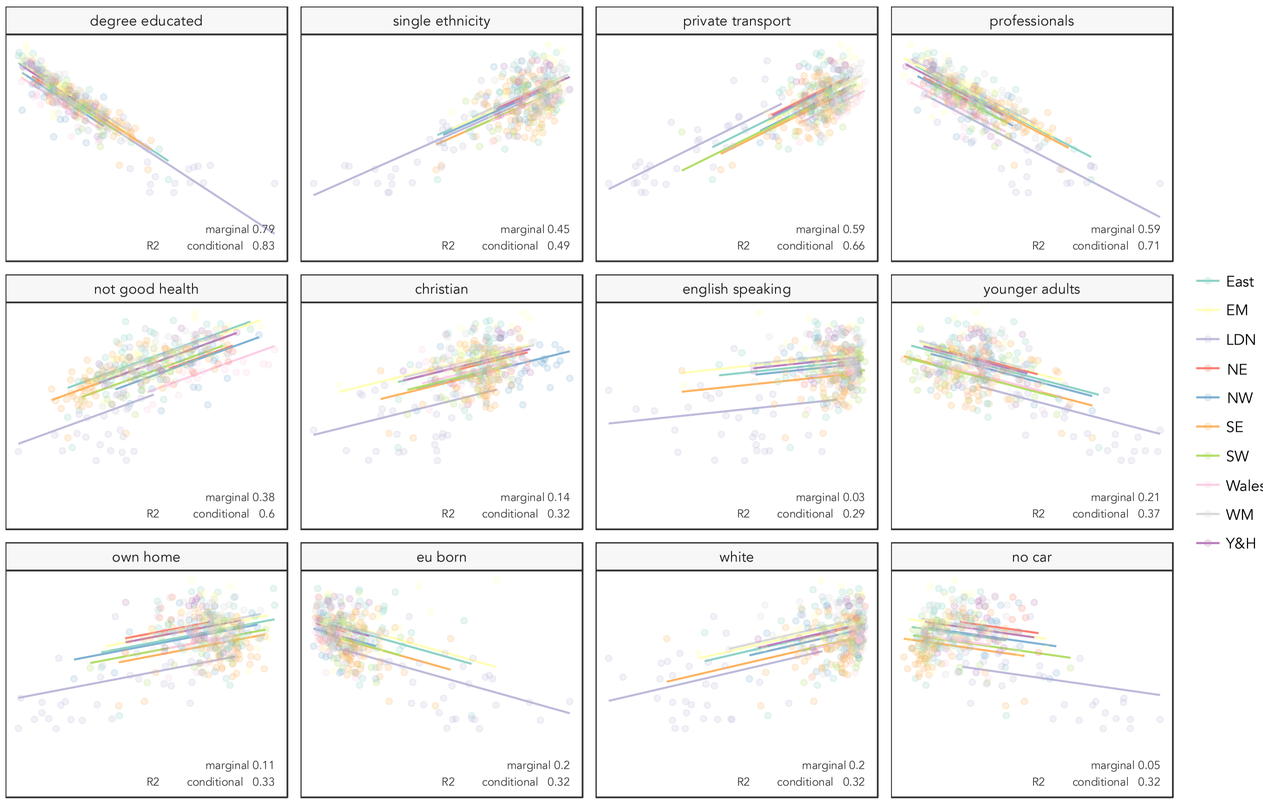


Figure 6: Varying intercept models fit for each explanatory variable. Dots represent observed values for each LA and are colored according to the region to which LAs belong; regression lines, offset by regional intercepts are also displayed. Estimates of pseudo-R^2 are in the bottom right [17]: marginal R^2 describes the proportion of variance explained by the fixed (non-regional) factors alone, conditional R^2 is the proportion of variance explained by both the fixed and random (varying-intercept) factors. A ColorBrewer 10-class Set3 qualitative scheme [13] is used.

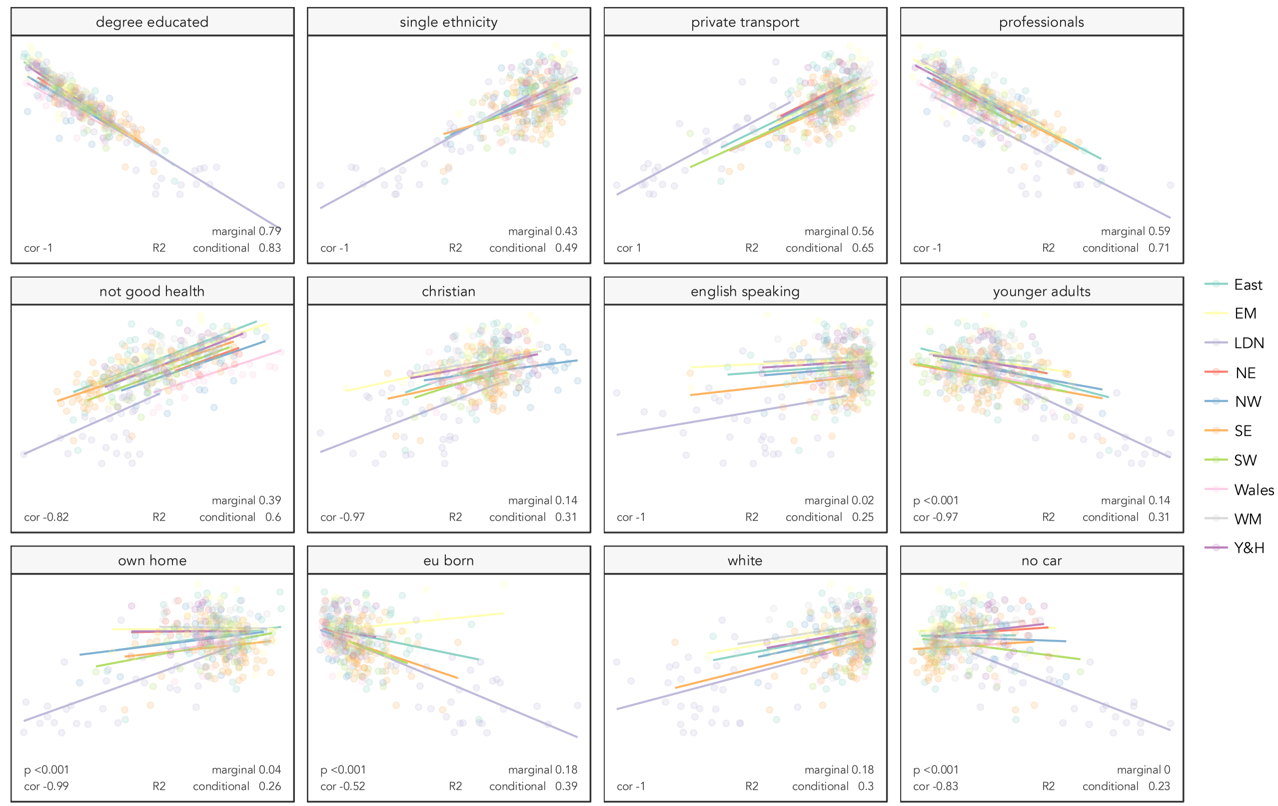


Figure 7: Varying slope models fit for each explanatory variable. Dots represent observed values for each LA and are colored according to the region to which LAs belong; regression lines, offset by slopes are also displayed. In the bottom left are results of a likelihood ratio test statistic comparing correlation values between slopes and intercepts. Estimates of pseudo-R^2 are in the bottom right [17]: marginal R^2 describes the proportion of variance explained by the fixed (non-regional) factors alone, conditional R^2 is the proportion of variance explained by both the fixed and random (varying-slope) factors. In the bottom left are results of a likelihood ratio test statistic comparing the varying slope with the varying intercept models and correlation values between slopes and intercepts in the slope models.

on these additional parameters. $p - values$ are generated by comparing this test statistic against a chi-squared distribution; where the resulting $p - value$ is ≤ 0.05, there is greater support for the claim that effects vary across region.

Visually inspecting Figures 7 and 6, we find further support for the claim that *degree-educated* is a global variable: the variable is not regionally structured (observed values are not clustered around regional regression lines) and the variable's effect on Leave (the regional slopes) does not vary. The same is true of *single ethnicity*, though here there is large variation also unaccounted for by between-LA context. The *professionals*, *not good health*, *Christian*, *English speaking*, *white* and *private transport* variables distribute differently from region-to-region and the addition of the varying intercept is found to improve model fit; however, after this adjustment, the effect on Leave is generally consistent between regions. Variables that exhibit both regional structure *and* effect are *EU-born*, *younger adults*, *no car* and *own home*. For the latter three variables a unique effect is identified for London and there is a negative correlation between slope and intercept: larger intercepts are associated with shallower slopes, thus these variables typically have the strongest effect on Leave in regions where support for Leave is comparatively weak. For the *EU-born* variable, the East Midlands is the only region where a positive slope is observed—here, this variable has the effect of increasing the Leave vote (coefficient for East Midlands of $+0.6$).

5 Discussion and conclusion

Consistent with existing area-based analyses [11, 12, 16], we find that Local Authority-level (LA) differences in EU referendum voting behavior can be related to differences in the demographic composition of those authorities. The most discriminating variables are those relating to *post-industrialization* and the *knowledge-economy*—those measuring (*degree-level*) education outcomes or distinguishing the relative prevalence of residents working in *professional* occupations. Variables linked to "traditional" values (*Christian*), metropolitan living (*no car*, *private transport to work*) and to a lesser extent material outcomes (*not good health*) help explain additional variation in voting preference. Given the geography of this variation, we echo the interpretation made in earlier studies [11, 12, 16] of the Leave vote being structured around parts of the country "left-behind" by globalization [5, 8].

A unique flavor to our analysis is around the use of local modelling techniques. We allow for the possibility that certain variables might be organized differently and have different *effects* in certain parts of the country than others. A large proportion of variation in the vote can be explained by a single variable (*degree-educated*) and our analysis confirms this as a "global" variable to the extent that its association with Leave does not change substantially across the country, allowing for regional distribution and effect in our multi-level models (as in Figures 6 and 7) does not noticeably improve model fit. For the secondary variables, there is greater evidence of local specificity. Variables relating (loosely) to nativist or traditional values (e.g., *Christian*) have an effect in the anticipated direction for more "provincial" regions of the country (e.g., not London or Scotland—Figure 4). For regions in the East Midlands, West Midlands and North West, a variable related to age and material outcome (*not good health*) is also identified. Slightly counter to expectation, for the North West, North East & Yorkshire and Scotland, variation in the *EU-born* variable has the effect of reinforcing the Leave vote. Studying local variation in *EU-born* in isolation, without conditioning on other demographics (geographically-weighted maps in Figure 5 and multi-level models in Figures 6 and 7), this variable is most strongly associated with Leave in Lincolnshire and parts of East Anglia: parts of the country where population change due to EU migration has perhaps been most keenly felt [1]. An important observation here, and to a lesser extent with other secondary variables, is that not only do the *levels* of individual variables change, but so too the interactions between variables—for the *EU-born* variable in particular the direction of relationship with Leave switches between positive and negative association.

Given the prevalence of the majority Leave vote (Figure 2), not all areas that voted Leave can be described as "left-behind." Our focus on locally-varying explanation offers some incremental insight here. Most obviously the fact that, within London and the south of England, additional variation is explained by variables that distinguish metropolitan contexts and "traditional" values (*Christian*), whereas further north, factors that describe material disadvantage (*not good health*) are also relevant. To further discriminate these subtleties, variables that capture themes such as historical Euroscepticism or exposure to post-2008 austerity measures would be worthy of investigation.

The limitations of area-based analyses such as that delineated in this study are well-rehearsed (e.g., [18]). Since we attend to associations between area-level demographics and area-level voting behavior, we can make claims around why certain authorities voted for Leave|Remain, but can claim nothing of the likely voting intentions of individuals. Also worth remembering is that regression modelling is a framework concerned entirely with variation: the mean and variance of one variable is assumed to represent some linear function of variation in other variables. In taking data aggregated to relatively large geographic areas (LAs), we smooth over important local variation and likely overstate the importance of social demographics at the expense of attitudinal explanations [14]. It is for this reason that we would argue for the spatial focus demonstrated in our study: if area-based approaches necessarily emphasize area-based explanations, then analyses that consider the spatial structuring of relationships and model outputs must be instructive. At a time when geographic differences in political preference appear to be widening—an obvious socio-geographic patterning has been observed around the recent US [9] and French [21] presidential elections—area-based approaches that are sensitive to spatial variation are especially prescient.

References

[1] ASTHANA, A. Immigration and the EU referendum: the angry, frustrated voice of the British public, June 2016. [Online; posted 20-June-2016].

[2] BATES, D., MÄCHLER, M., BOLKER, B., AND WALKER, S. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software 67*, 1 (2015), 1–48. doi:10.18637/jss.v067.i01.

[3] BRUNSDON, C., FORTHERINGHAM, M., AND CHARLTON, M. Geographically weighted summary statistics: a framework for localised exploratory data analysis. *Computers, Environment and Urban Systems 26* (2002), 501–524. doi:10.1016/S0198-9715(01)00009-6.

[4] BURN-MURDOCH, J. The demographics that drove Brexit, June 2016. [Online; posted 24-June-2016].

[5] COX, E. Leaving the North behind led to Brexit. Here's what has to happen next, June 2016. [Online; posted 24-June-2016].

[6] DORLING, D. Persistent North-South divides. In *The Economic Geography of the UK*, N. M. Coe and A. Jones, Eds. Sage, London, 2010, ch. 2, pp. 12–28. doi:10.4135/9781446269374.n2.

[7] FLORIDA, R. Toward the learning region. *Futures 27* (1995), 527–536. doi:10.1016/0016-3287(95)00021-N.

[8] FORD, R. Older 'left-behind' voters turned against a political class with values opposed to theirs, June 2016. [Online; posted 25-June-2016].

[9] GAMIO, L., AND KEATING, D. How Trump redrew the electoral map, from sea to shining sea, November 2016. [Online; posted 09-November-2016].

[10] GOODWIN, M., AND HEATH, O. Brexit vote explained: poverty, low skills and lack of opportunities, 2016. [Online].

[11] GOODWIN, M., AND HEATH, O. The 2016 Referendum, Brexit and the Left Behind: An Aggregate-level Analysis of the Result. *The Political Quarterly 87* (2016), 323–332. doi:10.1111/1467-923X.12285.

[12] HARRIS, R., AND CHARLTON, M. Voting out of the European Union: Exploring the geography of Leave. *Environment & Planning A 48* (2016), 2116–2128. doi:10.1177/0308518X16665844.

[13] HARROWER, M., AND BREWER, C. A. ColorBrewer.org: An online tool for selecting colour schemes for maps. *The Cartographic Journal 40*, 1 (2003), 27–37. doi:10.1179/000870403235002042.

[14] KAUFMANN. The 'Left Behind'?: reconciling individual and aggregate UK Independence Party voting. In *American Political Science Association meetings* (Philadelphia, USA, Sept. 2016).

[15] LU, B., HARRIS, P., CHARLTON, M., AND BRUNSDON, C. The GWmodel R package: further topics for exploring spatial heterogeneity using geographically weighted models. *Geo-spatial Information Science 17*, 2 (2014), 85–101. doi:10.1080/10095020.2014.917453.

[16] MANLEY, D., JONES, K., AND JOHNSTON, R. The geography of Brexit – What geography? Modelling and predicting the outcome across 380 local authorities. *Local Economy 32* (2017), 183–203. doi:10.1177/0269094217705248.

[17] NAKAGAWA, S., AND SCHIELZETH, H. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution 4* (2013), 133–142. doi:10.1111/j.2041-210x.2012.00261.x.

[18] ROSE, D. National and local forces in state politics: The implications of multi-level policy analysis. *American Political Science Review 67* (1973), 1162–1173. doi:10.2307/1956538.

[19] SIMON, N., FRIEDMAN, J., HASTIE, T., AND TIBSHIRANI, R. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *Journal of Statistical Software 39*, 5 (2011), 1–13. doi:10.18637/jss.v039.i05.

[20] THE GUARDIAN. EU referendum: full results and analysis, June 2016. [Online; posted 24-June-2016].

[21] THE NEW YORK TIMES. How the election split France, April 2017. [Online; posted 27-April-2017].

[22] TIBSHIRANI, R. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological) 58* (1996), 267–288. doi:10.1111/j.1467-9868.2011.00771.x.